

Welcome climate bloggers

A group of just nine climate scientists is trying to change the media coverage of their discipline. Thanks to an ongoing revolution in electronic news, they might just succeed.

For those excited about the journalistic potential of the Internet, this September was a seminal month. Much to its embarrassment, the US television network CBS was forced to retract a story about President George W. Bush's service in the armed forces. Crucially, the documents on which CBS based the story were first shown to be a fake not by a hot-shot Washington reporter, but by the authors of a series of blogs — online news diaries penned by unpaid commentators.

Blogs have been around since the late 1990s, but the political potential of this new media format only became truly clear during this year's US presidential race. Bloggers led some of the freshest debates, helped raise money for political parties and, most importantly from a research point of view, corrected mistakes made by other media outlets. At the same time, one more traditional website dedicated to this final purpose — www.factcheck.org — became compulsory reading for journalists who wanted to check the accuracy of claims made by the Bush and Kerry campaigns.

Can researchers make use of this new kind of communication? One group of climate scientists thinks so. The nine researchers (six in the United States, three in Europe) launched the RealClimate website — www.realclimate.org — earlier this month. As the News story on page 937 explains, the blog was motivated by the activities of some think-tanks, predominantly from the United States and often run using industry money. Ever since global warming became an issue of media interest, these groups have sought to play down its dangers. RealClimate, say the blog's founders, will provide rapid rebuttals of some of their more egregious statements.

Few would argue with the need to tackle attempts to distort science, but is a blog the best way to do it? The approach certainly has

its dangers. For example, many issues in climate science, such as the course of temperatures over previous millennia, are hotly debated by researchers. Some would argue that a rapid-rebuttal service, run with minimal peer review, can never hope to combat industry propaganda and properly represent this diversity of views.

Such criticisms are legitimate, but there is no reason that a prompt reply need be unbalanced. The researchers involved will, for example, have to work to ensure that they do not oversell their own opinions when commenting on research issues that divide scientists. Their goal is to provide solid scientific comment to journalists and other interested parties — and there is no reason to doubt that this can be achieved in this fashion.

What will happen if climate researchers do not take this risk? Industry lobbyists, as well as environmental organizations, will be free to distort science to fit their aims. These groups are expert at influencing the media, old and new. Press coverage of climate change is known to overly emphasize the views of the small minority of scientists who dispute the notion of man-made climate change (see *Nature* **431**, 4; 2004). Mainstream climate scientists need to combat this, and RealClimate is, in principle, an excellent way of doing so.

Similar exercises could aid researchers working on nanotechnology and transgenic crops, and other fields with a high media profile. *Nature* encourages scientists in these disciplines to consider setting up their own blogs, but also to monitor the progress of RealClimate. The site needs to balance speed with objectivity, readability and accuracy. That's no mean feat. Fail, and the blog will be dismissed as no more trustworthy than the myriad lobbying groups already writing on climate. But if the site's founders pull it off, they could change the coverage of climate change for the better. Good luck to them. ■

Wish list 2005

Nature hereby offers its readers some New Year's resolutions.

What would you like to see happen in 2005? We asked a diverse array of researchers — see page 942 for the results. But we decided to indulge ourselves too, and to impudently propose New Year's resolutions for our readers. Our wishes range from optimistic hopes for your even more responsible social behaviour to the more self-interested yearning that our own professional goals (and your enjoyment of *Nature*) could be made easier to achieve.

Good informatics. Resolve to visit a real library to browse paper journals and books at least once in 2005. And to remind yourself and students that available knowledge on a subject is not restricted to search returns on Google.

Good communication. Resolve to be brave and share your scientific knowledge with the general public, but resist the temptation to say any more than you know. Invite a journalist to lunch to get to know each other's crafts and needs. Offer to speak about your work in front of a school class — even a kindergarten class can be very rewarding and it's important for the young to meet real scientists. And in any

presentation, don't pack 50 Powerpoint slides into 10 minutes. Better still, avoid Powerpoint, and engage with people about what you do.

Good citizenship. Spend some of your free time contributing to a public good, such as the Drugs for Neglected Diseases Initiative. Speak out where dogma undermines good scientific advice. And, where you have the chance, resolve to take steps to keep creationism, 'intelligent design' and superstition out of science classes.

Good authorship. Resolve to never submit a paper to *Nature* until its overall goals and achievements can be understood by, at least, a scientist from another discipline, or even a non-scientist in your family. Always proofread your papers before submitting them to any journal. And avoid phrases such as 'Holy Grail' and 'Rosetta Stone' unless you really mean it.

Good leadership. For those leading research groups, resolve to (continue to) treat your graduate students and postdocs like people, not data slaves, to promote their work actively, and to put their career interests ahead of your own. And ensure that they, like you, have a great New Year. ■





Gloomy forecast
Climate conference
fails to generate
fresh actions
p936



Clouded view
Website launched to
combat global-
warming sceptics
p937



On the spot
Cosmetics firms
focus on Asian
skin problems
p938



Tooling up
Aid delivery seeks
to rebuild Iraq's
education system
p940

Activists and researchers rally behind AIDS drug for mothers

Erika Check, Washington

Scientists and patient advocates this week united to defend an HIV treatment against allegations that a key clinical trial of the drug was flawed.

The drug, nevirapine, is used to help prevent mother-to-baby transmission of HIV. But stories published by the Associated Press earlier this month re-ignited controversy in Africa over whether nevirapine is safe and effective, raising fears that many women there will stop taking the drug.

The allegations relate to a clinical trial — known as HIVNET 012 and funded by the US National Institutes of Health (NIH) — which began in 1997 in Uganda. In 1999 and in 2003, the scientists running the trial reported that nevirapine drastically cut mother-to-child transmission of HIV^{1,2}. The findings led to successful transmission prevention programmes around the world.

The Associated Press stories drew on internal communications between Edmund Tramont, head of the NIH's Division of AIDS, and his subordinates, who were concerned about standards of record keeping in the trial. Tramont, however, believed that a thorough review of the study had validated the conclusions that nevirapine was safe and effective, and stated this in a final report in March 2003.

Scientists and advocates are standing by Tramont's conclusion, pointing out that independent trials in South Africa, Malawi and Thailand³⁻⁵ have confirmed it. But nevirapine has long been a flash point in the AIDS epidemic, especially in South Africa (see *Nature* 430, 389; 2004), and the latest stories have revived vitriolic arguments about its use.

The African National Congress, the governing party in South Africa, issued an unsigned statement on 17 December, alleging that the drug was not proven to be safe but that the United States "was happy that the peoples of Africa should be used as guinea pigs". And on 16 December, the US political leader and civil-rights activist Jesse Jackson branded the NIH's continued support for the drug "a crime against humanity".

The NIH has asked the Institute of Medicine to perform an independent study of HIVNET 012, which will report in March.



Nevirapine is widely used in Africa to prevent the transmission of HIV from mother to child.

Officials at the NIH admit that the trial was not done perfectly, but they strongly support Tramont's decision to back the drug's use over the opinions of some of his subordinates.

"There certainly were elements of the study that could have been done better," says Clifford Lane, acting deputy director of the National Institute of Allergy and Infectious Diseases (NIAID). "But there's nothing that has in any way invalidated the conclusion that single-dose nevirapine is effective for reducing mother-to-child transmission."

Strong support

Academics also defended the conclusion. "Ed Tramont is a man of integrity and common sense, and he has probably done more than anybody at the NIH to improve the infrastructure for conducting these trials in developing countries," says Brooks Jackson, a pathologist at Johns Hopkins Medicine in Baltimore, who led HIVNET 012.

Brooks Jackson adds that he stands by the trial's results: "There's no question in my mind that single-dose nevirapine is safe and efficacious, and now that's been supported by several other studies independent of ours."

Even activists who often clash with the NIAID say that top scientists there correctly

interpreted evidence from the trial, despite some flaws in the study. "The data discrepancies don't alter the fundamental findings," says Gregg Gonslaves of Gay Men's Health Crisis, an AIDS lobby group in New York.

The furore over HIVNET 012 underscores the difficulties of doing clinical trials in developing countries. "You get criticized because some people feel you should provide the same exact standard of care as you do in the United States, and it's just not feasible," says Brooks Jackson. "On the other hand, you get criticized for doing regimens that are too difficult or cost too much or are not realistic, so you're sort of caught in the middle."

But activists are worried that the latest controversy is undermining efforts to prevent HIV transmission in Africa. "There are already mothers who are refusing to take nevirapine," says Arthur Ammann, a doctor and president of Global Strategies for HIV Prevention, a non-profit organization based in San Rafael, California. "This is the most successful therapy in the entire AIDS epidemic. It should not be attacked."

1. Guay, L. A. *et al. Lancet* 354, 795–802 (1999).
2. Jackson, J. B. *et al. Lancet* 362, 859–868 (2003).
3. Moodley, D. *et al. J. Infect. Dis.* 187, 725–735 (2003).
4. Taha, T. E. *et al. Lancet* 362, 1171–1177 (2003).
5. Lallamant, M. *et al. N. Engl. J. Med.* 351, 217–228 (2004).

Bush slips off the hook over funds for ocean management

Rex Dalton, San Diego

A top-level panel's demand for major reforms in US ocean management and research is getting the brush-off from the White House, say some of its members.

In a low-key announcement on 17 December, President George W. Bush said that he would set up an advisory committee to study the 212 recommendations of the Commission on Ocean Policy. These were published in September after a huge review of issues that ranged from fisheries conservation to coastal pollution.

But commission members are frustrated because Bush failed to promise substantial new funding. The commission's recommendations would have cost about \$1.5 billion in the first year, ramping up to nearly \$4 billion in annual spending after five years, according to its chairman, James Watkins, a former energy secretary.

"I'm a bit disappointed," says Andrew Rosenberg, a commission member and a fisheries scientist at the University of New Hampshire, Durham. The administration's response, which it was legally required to issue, "is very much a fig leaf", he says.

And Watkins himself says that the White House "fails to hit the nail on the head". Current funding for managing coastal waters "is inadequate", Watkins says, and the response does not specify by how much it should be increased. The oceans are "a precious resource", he adds, "and we don't have much time."

The commission was established by Congress in the final year of the Clinton administration in a bid to reinvigorate US ocean policy, but its members were chosen by the Bush administration. It took three years to research its findings, which are similar to, but milder than, those of a privately funded panel, the Pew Oceans Commission.

Environmental groups are even more critical of Bush's plan. "It's a big yawn; there is nothing there but a lot of hot air," says Gerry Leape, marine programme director at the National Environmental Trust. "It is a real missed opportunity."

But James Connaughton, a lawyer and chairman of the White House's Council on Environmental Quality, says that the administration will pursue a concrete response to the commission's recommendations. A list of research programmes worth supporting will be compiled by March 2005, he pledges. ■

US proves a wet blanket at international climate meeting

Amanda Haag, Buenos Aires

The latest global meeting on climate change wrapped up in Argentina this week, having made only modest steps towards cutting future greenhouse-gas emissions.

Delegates from many nations said that the United States and its allies, such as Saudi Arabia, thwarted progress at the tenth meeting of the Conference of Parties to the UN Framework Convention on Climate Change.

From 6 to 17 December, representatives of some 200 countries got together to discuss present and future climate-change negotiations. Russia's recent ratification of the Kyoto agreement on reducing greenhouse-gas emissions brings the protocol into force, so many delegates were keen to talk about what might happen after 2012, when Kyoto obligations expire. But the United States opposed such discussions. "We need to absorb and analyse lessons learned before committing to new actions," says Paula Dobriansky, head of the US delegation to the meeting.

During the conference, the United States strongly opposed the idea of using any seminars scheduled between now and next November to jump-start discussions about 2012 and beyond. Environmental groups attacked this position as deliberately obstructive. "I really think they've sunk to a new low here, by not only taking their own path but actively blocking other countries from pursuing the path they want to take," says Jeff Fiedler, climate policy specialist with the New York-based Natural Resources Defense Council.

"They never answered the question 'why are you objecting, why don't you just step aside' even though we asked them multiple times," says Debbie Reed, legislative director with the National Environmental Trust, a Washington-based environmental coalition.

Delegates haggled until sunrise on the last day of the conference over the wording that will regulate the seminars. In the end, it was agreed that just one seminar would take place before the next annual meeting to "promote an informal exchange of information" and to "continue to develop" appropriate responses to climate change. This at least leaves an opening for discussions about policy after 2012, says Elliot Diringer, director of international strategies with the Virginia-based Pew Center on Global Climate Change.

In future, the United States may have less sway over such issues. The 2005 conference will be split into two sections, one stream of which will be for Kyoto parties only: US delegates may not be able to take part in these sessions.

Saudi Arabia also caused dissent at the meeting, by asking for money from the 'adaptation fund' to offset the economic losses it will suffer when petroleum exports are reduced. By 2010, the country expects lost fossil-fuel exports to cost it billions of dollars annually. But the adaptation fund is meant predominantly to compensate developing countries and vulnerable island nations.

Although Saudi Arabia always raises the issue of compensation, this year's suggestion that it be provided by the adaptation fund was especially contentious, says Reed. "We're having a hard time raising money for developing countries," says Reed. "So for a country like Saudi Arabia to demand compensation for lost sales of oil — it's worse than ironic."

Despite these difficulties, the meeting struck some hopeful notes. The European Union and other nations renewed a pledge to deposit \$420 million annually, beginning in 2005, to fund developing countries' efforts. And, following Russia's lead, Indonesia and Nigeria have ratified the Kyoto Protocol. ■



Raining on the parade: the United States put a dampener on the tenth UN climate-change convention.

D. LUNA/AP

Climatologists get real over global warming

Jim Giles, London

It seemed like another public relations setback for climate science. In *State of Fear*, the latest blockbuster from best-selling author Michael Crichton, in-depth discussion of climate change rubs shoulders with car chases and shoot-outs. The book echoes the stance of climate sceptics, rubbing predictions that global temperatures are set to rise.

But one group of climate researchers refused to take such talk lying down. Within a week of the book's publication on 6 December, a detailed critique of its contents was posted on a new website targeted at journalists and the public. The site, RealClimate, which put up its first posting on 1 December, is the brainchild of a group of respected climate scientists who are determined to tackle what they see as poor media coverage of their field.

"We're trying to be a quick-reaction squad," says Ray Bradley, director of the Climate System Research Center at the University of Massachusetts in Amherst. Bradley is one of nine scientists, based in Europe and the United States, who will initially contribute to the site. "We all felt there was far too much disinformation propagated through the media," he says.

The site aims to provide responses to and context for press coverage of climate research. The founders say that they will stick to topics on physical science and steer clear of political and economic issues.

The aggressive posture of the site, which contained two postings disputing Crichton's opinions, together with many follow-up comments, when *Nature* went to press, was welcomed by many researchers.

But some sound a note of caution. They

say that the site risks presenting a "party line" on climate science that could be seen as attempting to limit debate. The site's critics also fear that, by rushing out responses rather than going through peer review, the site blurs the line between science and advocacy.

Gavin Schmidt, a climate modeller at NASA's Goddard Institute for Space Studies in New York who helped to launch the site, concedes that it is "in some senses" a vehicle for advocacy, rather than science. But he says it is needed to counter pressure groups funded by the US energy industry that deny that global warming is happening and is being caused by greenhouse-gas emissions. Such groups are "truly abusing scientific results", according to Schmidt.

Some of the researchers who agree with Schmidt's observation are not sure that



Heated debate: Michael Crichton's take on climate change has prompted online rebuttals.

projects such as RealClimate are the best response. "You could get the same names criticizing every time," warns Mick Hulme, director of the Tyndall Centre for Climate Change Research based in Norwich, UK. "That would give the impression that there is a party line."

Critics also suggest that the project may struggle to accommodate respected climate scientists who dissent from aspects of mainstream thinking on climate change. One researcher mentioned in this regard was Richard Lindzen, a meteorologist at the Massachusetts Institute of Technology in Cambridge, who has questioned the current scientific consensus over the impact of carbon dioxide emissions on climate. The possibility of posting his views seems to divide the RealClimate team. Schmidt said he was unsure whether Lindzen would be allowed to post, but Bradley said that he should be.

And Rowan Sutton, a climate modeller at the University of Reading, UK, points out that conventional peer review is incompatible with the site's need for rapid response. "It's a dilemma for them," he says. Schmidt counters that postings are not academic papers and so do not need full peer review. Comments are instead e-mailed to researchers contributing to the site, and their suggestions are incorporated before the piece is uploaded.

In the short term, simply keeping the site up and running is the team's greatest concern. Schmidt says they are getting more external comments than they had expected and have already had to relax the rules on how such material is vetted before being added to the site. "It's a huge time sink," agrees Bradley. "That's my only fear."

♦ www.realclimate.org

J. COOPER/AP

Funding review sparks fears for Canadian science

David Spurgeon, Montreal

Canadian scientists are anxiously awaiting the results of a government spending review, amid fears that it could reverse the recent spurt in federal support for research.

The review, which is scheduled to finish in January, aims to cut the overall federal budget by Can\$1 billion (US\$820 million) for 2005 and to reallocate Can\$12 billion over five years from 'low-' to 'high-priority' areas.

But the move has prompted unease among scientists because their funding agencies, in common with other government departments, have been asked to come up with options for cutting 5% from their budgets to aid the reallocation plan.

"There is certainly anxiety" about the process, says John Challis, vice-president of research at the University of Toronto, the

nation's leading research university.

"Word of the 5% cut is now getting out into the university community and to the department chairs and deans, so I'm starting to get questions from them." Cuts of this magnitude "would have a deleterious effect on Canadian science", he adds.

Canada's research councils are set to offer Can\$94 million towards the 2005 budget cut from funds that weren't used this year for the Canada Research Chairs programme — although they hope that they will have to make such a contribution for one year only. They remain unsure whether wider cuts will be imposed on their budgets, which have grown quite quickly in recent years.

But in the past few weeks there have been hints that research agencies could be exempted from large-scale cuts. Tom

Brzustowski, president of the Natural Sciences and Engineering Research Council (NSERC), Canada's main granting body for non-medical university research, says that he was worried earlier on in the process. However, he says that he was "very encouraged" after David Emerson, the industry minister and a member of the panel reviewing the budget, reaffirmed the government's commitment to university research at a parliamentary committee hearing on 25 November.

And Nigel Lloyd, executive vice-president of the NSERC, says that "some assurances" have been received "from senior-placed people" that the councils will be exempted from the cuts. Nevertheless, he adds that he'll remain concerned until the review outcome is announced.

Spotlight turns on cosmetics for Asian skin

Ichiko Fuyuno, Tokyo

Cosmetics companies and scientists around the globe are expanding their efforts to understand Asian skin problems, a symposium was told earlier this month.

Many skin treatments have been developed for Caucasian women, but women in east Asia have different skin problems, say specialists. These include dark melanin spots, which result mainly from exposure to the Sun's ultraviolet rays.

On 11 December, 16 researchers from France and several Asian countries addressed a symposium in Paris organized by Moët Hennessy Louis Vuitton (LVMH), a Paris-based company specializing in luxury brands, including cosmetics. The researchers described their latest findings on skin ageing, the skin's reaction to stress, and those melanin spots.

"In the past, researchers just tried to get rid of spots that appeared on the skin," says Hachiro Tagami, a dermatologist at Japan's Tohoku University in Sendai. "But now they are trying to find out why spots develop."

In general, researchers say, dark spots appear when the activity of melanocytes is disrupted, usually by excessive exposure to ultraviolet light. Melanocytes are cells that produce the pigment melanin, and their activity levels help to determine the skin's condition, as well as its colour.



Beauty spots: the skin of Asian and Caucasian women differs in its sensitivity to the Sun's ultraviolet rays.

Asian women find that melanin pigments develop in their facial skin at earlier ages than they occur in Caucasian women, say researchers, and that the activity of their melanocytes seems to be more sensitive to stress. But Asian skin retains more of its tension and is less likely to wrinkle.

Researchers are now gaining insight into the genetic and biochemical differences that might explain these contrasts.

In a collaboration with Bordeaux University, for example, workers at LVMH have come up with new data in the search to explain melanin spot formation. Beneath such spots, the melanocytes, which are usually found in the epidermis, seem to sink into the dermis by disrupting the thin structure between the layers known as the dermo-epidermal junction (M. Cario-Andre *et al.* *J. Cutan. Pathol.* **31**, 441–447; 2004).

LVMH says it is developing these findings to launch a skincare range for east Asia under its Christian Dior brand. Meanwhile, cosmetics researchers at Procter & Gamble claim to have found a method of controlling the transfer of melanin from melanocytes, using a substance called niacinamide. (T. Hakozaiki *et al.* *Br. J. Dermatol.* **147**, 20–31; 2002).

Regulators will be keeping a close eye on the safety of these skin treatments, however. Last year, Japan's health ministry ordered 12 Japanese and foreign cosmetics companies to stop selling products containing kojic acid, which is believed to reduce melanin levels in the skin. The agent's cosmetic use was approved in 1988, but animal tests have raised fears that it may cause cancer. ■

Nano names go on Internet sale in hope of mega profit

Jim Giles

For sale: hundreds of Internet domain names, all starting with 'nano'. Vendor prefers to sell collection as a single unit — would suit large funding organization.

This curious offer is being made by Steve Sobol, a US environmental consultant and Internet entrepreneur. Over the past two years, Sobol's company ESH Sciences has snapped up more than 900 domain names in the hope of selling the collection to an organization such as the US National Science Foundation (NSF).

Trading in domain names can be big business. Speculators can register names for as little as US\$7 a year and, if they choose well, resell for a considerable profit. One of this year's biggest transactions involved the domain name *creditcards.com*, which sold for \$2.75 million. Such returns entice entrepreneurs to bag large numbers of names in the hope that one

will turn out to be hot online property.

"Domains with 'nano' in them haven't appeared in our charts of most costly names," says Ron Jackson, editor of *Domain Names Journal*, an online magazine. "But nanotechnology has huge potential. That's why it's worth locking up the domains for future use."

Sobol, who describes the venture as providing a "very good public service" with "some greed in there too", declines to say how much his collection is on sale for. But he points out that nanorobotics.com is currently available from another source for \$470, and says that nanocolor.com sold earlier this year for \$3,300.

So far, Sobol has not had any offers for his collection. He says he contacted the NSF in June 2003 about the names but received no reply. When he tried to attract attention to his venture by advertising it on an online science-discussion group, he received only

one reply: he was accused of 'cyber-squatting' — a derogatory term for the practice of buying up domain names that should be owned by somebody else.

A partial list of Sobol's domains, available on his website *naknow.com*, includes such oddities as 'nanosociology' and 'nanocrimelab'. He also holds at least one name that corresponds to an actual research project: 'Nanocosmos' is a collaboration into semiconductor research funded by the European Union. Project leaders could choose to take Sobol to task over the name, and apply to a tribunal to ask that the site be handed over at cost. But it is unclear whether such a move would be successful — or worth the effort. ■

For more news and analysis go to
news @ nature.com
www.nature.com/news

US puts end to block on papers from sanctioned states

Washington Months of confusion ended last week when the US Department of the Treasury issued licences that permit US publishers to accept and print scholarly papers with authors from Cuba, Iran and Sudan. Some publishers had stopped doing so over concerns that they could fall foul of US trade sanctions on these countries, which are deemed to pose an “unusual and extraordinary threat” to US national security.

The department's Office of Foreign Assets Control (OFAC) initially ruled that publishing papers from these countries required a special licence (see *Nature* **427**, 663; 2004). Publishers who failed to obtain one would be violating the trade embargo and be liable to heavy fines, it said.

The Institute of Electrical and Electronics Engineers was exempted from the requirement in April. But four groups representing academic publishers decided to sue the treasury on the grounds that the requirement is unconstitutional.

Allan Adler, vice-president for legal and governmental affairs at the Association of American Publishers, one of the groups filing the suit, says that the department's

decision is a positive one. He and the other publishing groups will now review whether to proceed with their lawsuit.

Books and lab supplies head for embattled Iraq

London A convoy carrying US\$4.6 million worth of laboratory equipment and \$1 million worth of academic books is bound for higher-education institutions in Iraq. The consignment — organized by the International Fund for Higher Education in Iraq, the United Nations Education and Scientific and Cultural Organization (UNESCO) and the Doha-based Qatar



In need of help: Iraqi labs such as this one at al-Mustansiriyah University are ill equipped.

Foundation — is a bid to help rebuild Iraq's battered higher-education system.

After local consultation, pharmaceutical, medical and engineering labs were chosen to benefit, although their identities are being kept secret for security reasons. The delivery will be distributed under the escort of the Iraqi authorities. “Iraq's universities and technical institutes are in a dramatic situation,” says Koichiro Matsuura, director-general of UNESCO. “They have suffered severely from war damage and neglect.”

Bubble-based fusion bursts onto the scene

Washington A company in California is launching an experimental power reactor based on ‘bubble fusion’, despite reservations within the scientific community over whether the effect exists.

Impulse Devices in Grass Valley hopes to sell its sonofusion research reactors for about US\$250,000. It claims they use ultrasound to generate bubbles in ‘heavy’ water, made up of hydrogen's heavier isotope deuterium. The bubbles can be imploded rapidly, generating a high temperature that allows deuterium nuclei to undergo fusion reactions, it says. “The technology could produce enough energy for electricity production in ten years,” claims

Mark Ludwig, chief executive of Impulse.

But many scientists are not convinced. Researchers at Oak Ridge National Laboratory in Tennessee claimed to have achieved fusion with a similar technique in 2002. But an internal review by other Oak Ridge scientists questioned the group's results, and the work remains in limbo (see *Nature* 416, 7; 2002).

Hungarian head named for launch of 'European CDC'

Stockholm Zsuzsanna Jakab, secretary of state at Hungary's ministry of health, was last week nominated as the inaugural director of the European Centre for Disease Prevention and Control (ECDC).

The new European Union agency, which is based in Stockholm, Sweden, will collect, evaluate and disseminate information about infectious diseases throughout Europe. At the moment this job is done by individual member states.

The ECDC is modelled on the US Centers for Disease Control and Prevention in Atlanta, Georgia. But it faces a difficult start because of its limited budget and its lack of associated laboratories and regulatory power (see *Nature* 431, 507–508; 2004). Jakab's appointment is expected to be confirmed by the European Parliament early next year.

Ivory flute brings ice age's cool music to life

Munich Archaeologists in Germany last week unveiled what they claim is one of the world's oldest, and for its age most sophisticated, musical instruments.

Nicholas Conard, an archaeologist at the University of Tübingen, and Friedrich Seeberger, an expert in ancient musical instruments from Radolfzell, described an ivory flute that they rebuilt from fragments found in a cave in mountains near Ulm in southern Germany. It seems to have been carved by the cave's ice-age dwellers from a mammoth's tusk. They



estimate from dating the surrounding deposits that the flute is more than 30,000 years old.

Seeberger has made a wooden copy of the flute that he says can make harmonic and versatile music, apparently following the pentatonic scale common in Asia, rather than the diatonic tones familiar to present-day Europeans.

Painkiller trial halted over heart scare

Washington A US clinical trial of the prescription painkiller Celebrex was halted on 17 December after people taking the drug showed an increased risk of heart problems.

The trial was investigating whether Celebrex, one of a class of drugs known as COX-2 inhibitors, could help to prevent colorectal cancer. Pfizer, the company that makes the drug, says it has not seen an increased risk of heart trouble in its own

trials of the painkiller. It has so far not removed the drug from sale, but says it has stopped actively advertising it.

COX-2 inhibitors made the headlines in September, when Merck's Vioxx was removed from sale after it was linked to an increased risk of heart attacks and stroke.

The Food and Drug Administration is suggesting that doctors avoid prescribing COX-2 inhibitors, although it has yet to issue any formal guidance. The National Institutes of Health will now review all of its 40 or so clinical trials involving the drugs.

All I want for 2005 is...

2004 was a year dominated by news of war, disease and dark threats of bioterrorism and nuclear proliferation.

But for the scientific community there were signs of progress and hope. Researchers continued to interrogate nature and produce intriguing results, ranging from the genetic sequence of the rat to the surprise discovery of a previously unknown and diminutive species of humankind. And scientists exasperated by the reluctance of policy-makers to take research seriously could draw some comfort, at least, from Russia's ratification of the Kyoto Protocol on Climate Change. In the following pages, *Nature* highlights some of the events that excited the world of science over the past 12 months.

And as the calendar pages flip over to a new year, we've also been making some enquiries into what scientists think — and hope — is in store next. Given no constraints, we asked, what would you wish for in 2005? What follows is a sample of the responses our reporters got from researchers and policy-makers across all fields of study, along with summaries of the events that framed these aspirations. From the development of effective therapies against HIV and malaria, to the detection of life elsewhere in the Universe, to the ability to do research without the intrusion of politics or bureaucracy, the community's wish list provides an insightful — and refreshingly optimistic — glimpse into the future.

“Open access to America from abroad.”

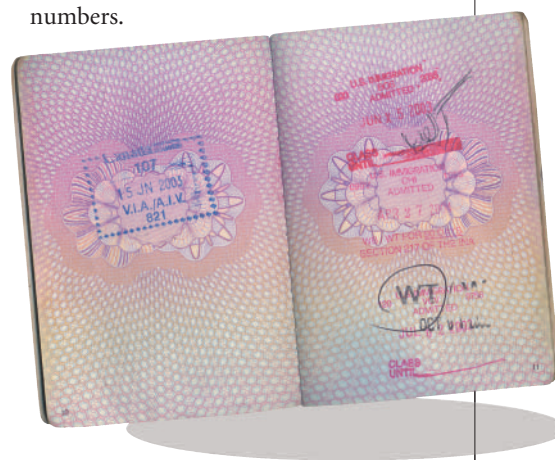
David Baltimore
President, California Institute of Technology

Students, scholars and at least one Nobel prizewinner have had trouble getting into the United States this year, thanks to immigration rules that have grown ever tighter since the terrorist attacks of 11 September 2001. Zhores Alferov, who won the 2000 Nobel Prize in Physics for his work on semiconductors, stormed out of the US consulate in St Petersburg, Russia, this September without his visa after being grilled about the nature of his work.

The situation has improved somewhat since heavy restrictions were introduced shortly after the attacks. Most background checks are now completed within 30 days — half the time it used to take — and are valid for up to a year, making it easier for foreign scientists working in the United States to travel home for holidays and family events. But it remains unclear whether these reforms will be enough to stop foreigners from spurning US academic institutions. This year's omens were not good: for the first time in more than three decades, the number of international students enrolling in the United States fell.

This is a trend that Baltimore finds deeply disturbing. With Europe and Asia becoming increasingly competitive, he says, the United States no longer has a firm lead in research. One of the reasons for this is that it's now so much harder to come to the United States to study, he argues.

Meanwhile, other nations are making the most of the United States' tough new rules. The number of foreign science students enrolled at universities in Australia has shot up by 32% since 2001. Asian students seem to be flocking to Britain too: the University of Cambridge, for example, has seen a surge in students from China. And China, in turn, is welcoming students from neighbouring countries such as Indonesia in record numbers.



“A malaria vaccine that really works and is cheap enough for African kids to afford.”

Gustav Nossal
Immunologist, University of Melbourne, Australia
A trial vaccine, known as RTS,S/AS02A, was shown this year to shield some children from malaria: the first real success in the field. Much more work needs to be done to achieve full protection, and to make the jabs affordable. But more trials are under way.

“I’d wish for two burning plasma experiments in the world, instead of just one.”

Gerald Navratil
Plasma physicist, Columbia University, New York

For plasma physics, 2004 was characterized by the constant dispute over where to build ITER, an international experimental reactor aimed at producing power from the fusion of hydrogen atoms. The community is united in its desire to see the project go ahead. But the six partners — Russia, China, South Korea, Japan, the United States and the European Union — are currently deadlocked over whether the reactor should be located in France or Japan.

The stalemate has led the Europeans to decide that, if necessary, they will go it alone. That could be okay, laughs Navratil, if it means that both the European and Japanese consortiums each build a machine. “Obviously we’d like to have at least one ITER,” he says; but two would be even better. Observers, however, would be extremely surprised if this wish actually came true.

“I wish for a cataclysmic rearrangement of the tectonic plates — or alternatively some creative legislative gerrymandering — so that the San Andreas Fault line ends up just west of Boston, Massachusetts.”

George Daley
Stem-cell scientist, Harvard University

It seems as though nothing short of serious seismic upheaval will be enough to get researchers on the US east coast the money they want to study human embryonic stem cells.

Federal funds for such work remain limited to a few dozen cell lines. But on the day of George W. Bush’s re-election,



a referendum in California backed an initiative to plough \$3 billion of state funds into the field, turning a lot of researchers farther east green with stem-cell envy.

Those outside California aren’t completely bereft. In April, Harvard University announced the creation of a stem-cell institute in and around Boston involving 100 researchers and funded with millions of dollars of private money. Three months later, the governor of New Jersey signed legislation to spend \$9.5 million on stem-cell research.

The stage is now set for regulatory battles between the conservative federal government and those states using public money to pursue embryonic stem-cell research. A bill to ban ‘therapeutic cloning’ — which would use genetic material from a cloned embryo of the patient to make new cells for a potential transplant, for example — that has languished since 2001 may pass next year, thanks to the newly enlarged Republican majority in the Senate. Even if this bill doesn’t pass, limitations on stem-cell research could also be tacked on to unrelated legislation and end up as law if supporters of the research fail to muster the political muscle to stop them.

“I wish to see tranquillity, security and freedom of thought granted for scientists and researchers in parts of the world suffering political turmoil.”

Radwan Barakat
Plant scientist, Hebron University,
Palestinian Authority

©2004 Nature Publishing Group

Highlights

Bring in the clones

Scientists in South Korea successfully cloned 30 human embryos and extracted stem cells from them. Others have claimed to do the same before: a US company announced in 2001 that it had produced a short-lived cloned embryo, for example. But evidence for such reports has been scarce until now. It took 242 eggs from 16 women to make these few clones. If the process can be streamlined, some hope it could one day be used to provide replacement cells tailor-made for any patient.

The hobbit

The skeleton of a tiny hominid was unearthed in Indonesia, providing evidence for a previously unknown branch of human evolution that lived just 18,000 years ago — a startlingly contemporary date. The one-metre-high species, called *Homo floresiensis* after the island it was found on, and nicknamed ‘hobbit’ by its finders, has made researchers wonder what other creatures might be out there waiting to be found.

Tracking electrons...

A magnetic microscope was tuned to detect the tiny signal of a single electron’s spin. The device — a minuscule cantilever with a magnetic tip — wobbles in the presence of the magnetic field created as a single electron spins. This wobble is in turn tracked by a laser. Although the principle is simple, it took the research team 12 years to attain the sensitivity needed to detect a single electron. The researchers say that the same technique should be able to detect an electron’s spin orientation, aiding attempts to produce quantum computers.

...and taking their picture

A snapshot was taken of an electron orbital, by blasting a nitrogen molecule with laser pulses. The light emitted after these kicks from the laser reveals an image of the space where the molecule’s electrons reside. The shapes and sizes of electron orbitals have been determined in the past through experiment and theoretical calculation, but this is the first time their picture has been taken. The same technique should some day let researchers watch electrons as they take part in chemical reactions.

Rat tales

The Brown Norway rat became the third mammal to have its genome sequenced, joining mice and humans. The rat is a model of choice for many studies in physiology and pharmacology, and is used to investigate everything from cardiovascular disease to space motion sickness. Researchers pored over the animal’s 25,000 genes to enlighten this work. Next year should see several more mammals join the club of sequenced creatures, including the chimpanzee and the dog.

“A definitive cure in gene therapy for some sort of routine disorder that’s applicable to a large number of other diseases.”

Mark Kay
Gene-therapy researcher, Stanford University

This year, the gene-therapy field began regrouping after a difficult period — and received a shot in the arm from a young technology called RNA interference.

In France, authorities allowed Alain Fischer of the Necker Hospital in Paris to restart a gene-therapy trial that had been on hold for almost two years. The trial uses gene therapy to cure children who have the fatal condition X-linked severe combined immunodeficiency disease (SCID), which leaves sufferers unable to fight off infections. But it and other SCID gene-therapy trials around the world have been on hold since last January because Fischer’s treatment caused cancer in two out of eleven children.

Now, regulatory authorities in France and elsewhere have decided that the SCID trials can resume, because the alternative — bone-marrow transplants — isn’t always successful. In the United States, the Food and Drug Administration (FDA) has likewise decided to allow at least one trial to go ahead.

But will gene therapy prove successful? Enter RNA interference, a technique that takes advantage of natural human defence mechanisms and that many researchers think could deliver the



first full cure in molecular medicine. Biotechnology companies seem to agree; this year, two of them — Sirna Therapeutics in Boulder, Colorado, and Acuity Pharmaceuticals in Philadelphia, Pennsylvania — filed applications with the FDA to begin clinical trials using RNA interference to treat macular degeneration, a progressive eye disease.

They are likely to be joined next year by Alnylam Pharmaceuticals of Cambridge, Massachusetts. Researchers at this firm have already demonstrated that RNA interference can be used to cut cholesterol in mice. A cholesterol-lowering treatment would be blockbuster for RNA interference, but that is still years away.

In the immediate future, look for more clinical trials in 2005 — including some using RNA interference to combat hepatitis C or HIV.

But although some scientists were grumbling about the limits of today’s transportation, others were working on a project that they think could revolutionize tomorrow’s. Proving that space is accessible to your average billionaire as well as to space agencies, aerospace designer Burt Rutan and Microsoft co-founder Paul Allen launched the first private rocket to the outskirts of suborbital space and scooped the US\$10-million X prize in the process.

Rocket enthusiasts celebrated the achievement as the dawn of a new era of space tourism. But sceptics said that private space travel is unlikely to take off until engineers conquer the much harder feat of getting tourists into orbit. Virgin Galactic expects to begin commercial flights as early as 2007, with seats going for about \$200,000 a pop. That probably falls outside the reach of most researchers — but some, at least, hope that the technology will one day find a use in faster-than-Concorde intercontinental travel.

“My wish? ET: call me.”

Louis Friedman
Executive director, Planetary Society

For Friedman, who heads a large space-advocacy group, there is no question about the major goal of space exploration: it is to find life. To this end the Planetary Society strongly supports SETI — the search for extraterrestrial intelligence that scans the skies for signs of communication.

In contrast, NASA now seems relatively unsure of its goals. Should it finish building the International Space Station? Stick with robots for most voyages, or push for piloted missions?

The Bush administration tried to create some focus this year by declaring that NASA would put astronauts back on the Moon by 2020, and then head for Mars. Initial funding for the president’s ‘Vision for Space Exploration’ was passed by Congress in November despite reservations from many lawmakers and scientists. It may provide a focus for the space programme, but it doesn’t seem to be a wildly popular one.

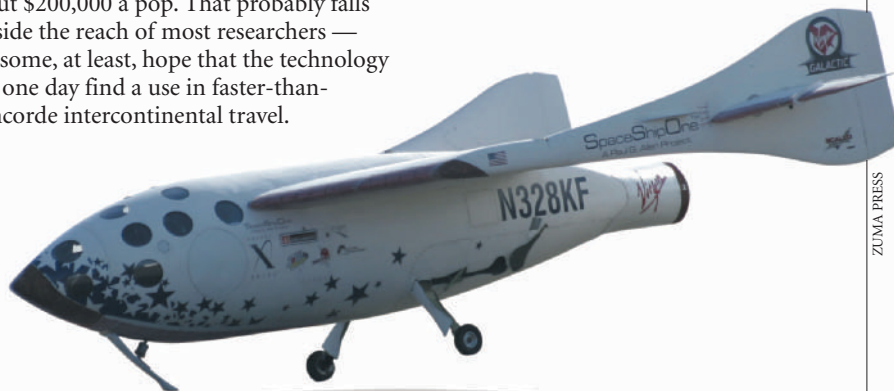
Some US scientists worry that if the tide shifts back to an expensive astronaut programme, it will detract from pure research without much obvious benefit. Missions already launched won’t feel the squeeze — including Messenger, which is due to arrive at Mercury in 2011, and the Cassini craft, which should release a probe down to one of Saturn’s moons, Titan, in January. But plans for spacecraft to study black holes and dark matter have been put on hold, conceivably delaying our discovery of wormholes and pockets of alien life in far reaches of the Universe. Unless, of course, ET calls us first.

“A cure for jet lag — or super-fast flight. And fool-proof, fast, invisible airport screening technology. Actually, I’d settle for super-comfortable flight.”

George Daley
Stem-cell scientist, Harvard University

If there’s one thing scientists agree on, it’s that they spend too much time crunched up in economy class and not enough in the lab — and travel only seemed to take longer in 2004.

The United States, for one, lengthened queues when it phased in the largest biometric scheme yet deployed at national borders, demanding that foreign visitors give electronic fingerprints for checking against a database of undesirables. Next year, many countries are expected to introduce passports that encode biometric information about their owners on microchips. This will either shorten or lengthen queues, depending on your faith in technology.



"I'd wish to find out if there is life on Mars; perhaps martians can be detected by a whiff of their farts."

Roger Buick
Geologist, University of Washington, Seattle

If you're looking for extraterrestrial life, then Mars is a great place to start. Tantalizing discoveries this year meant that the possibility of finding microbial life — ancient or contemporary — on the red planet once again came to the fore.

Earth-based telescopes and Mars Express — a European mission sent to orbit the planet — detected the presence of methane in the martian atmosphere. As methane is a short-lived gas, researchers say that this must have been produced within the past 300 years or so. With no known active volcanoes on the planet to generate the gas, this has left researchers "twitching and excited" about the possibility of contemporary microbes as the source, says Buick.

Is there a way to find out if the methane comes from life? Buick suggests that we now look for traces of hydrogen sulphide — another gas commonly produced by biological activity. If sufficient volumes of hydrogen sulphide are coupled with the leaking methane, it would suggest that subsurface life is producing the gases, he says. Mars Express did see hints of hydrogen sulphide, but these measurements have not been confirmed.

NASA's martian rover, Opportunity, also found clues to add to the growing body of evidence that Mars once held liquid water: marks in rock that looked as though they had made by ripples; sulphate and other deposits that seemed to have been left when a briny pool evaporated; and tiny spherical rocks that probably formed as minerals precipitated out of water bubbles. Sadly Europe's equivalent lander, Beagle 2 — the only one explicitly designed to look for signs of life — didn't survive the trip to the planet's surface.

"For the French government to take on board scientists' proposals for the future of research, so that France once again becomes an attractive destination for young scientists."

Alain Trautmann, cell biologist and leader of Save Research, an unprecedented scientific revolt against French government policies and science funding

"For legal enforcement of the Hippocratic oath 'first, do no harm' to ensure that all physicians and researchers are held accountable if they violate ethical standards."

Vera Sharav
President, Alliance for Human Research Protection, New York.

Every year has its share of people who lie, steal, cheat or fall prey to subtler ethical slip-ups in the lab.

Although 2004 wasn't the worst of recent years, an array of scientists were nevertheless accused of plagiarism, fraud and other misbehaviour. Even the editors of journals confessed to the occasional ethical slip-up in their publishing practices — such as asking authors to add specific references to their papers to boost the journal's impact factor. As a result, one association of medical journals, at least, has drawn up a code of good practice for themselves to keep things in line, which came into force this month.

The pharmaceutical industry came under attack when GlaxoSmithKline (GSK) was accused of suppressing results of clinical trials that suggested some antidepressants could increase the risk of suicidal behaviours in children. Rules were changed in the United States to ensure that these drugs were labelled with a warning, and that more data would generally be made available for public scrutiny. Twelve leading international medical journals decided that companies would have to register details of clinical trials in a public database if they want to have their results published. This, they hope, will redress the fact that only trials with positive results tend to be published or aired in public. GSK also promised to put summaries of its clinical-trial data for marketed drugs online for free — and has begun to do so.

The physical sciences suffered some unusual problems as well: staff were fired from Los Alamos National Laboratory in New Mexico after some classified computer-storage devices disappeared, forcing the lab to shut down for weeks for a security review.

And one of this year's science highlights — the successful cloning of human embryos in South Korea — was clouded by suspicions that one of the lab's researchers was the source of some of the eggs; usually considered to be ethically unacceptable. Lab chief Woo Suk Hwang at Seoul National Laboratory put his work on hold after the accusations hit. A national bioethics law that comes into effect in January 2005 may help to sort out future issues.

Highlights

Mum's the word

Mouse eggs were persuaded to grow into apparently healthy mice without being fertilized by sperm, making for the first birth of a mammal without contribution from a father. The success doesn't make men irrelevant — the genetic manipulations used by the team are for now, at least, technically and ethically infeasible in humans: the experiments produced far more dead and defective baby mice than normal ones.

As old as ice

Results have started to pour in from a core of Antarctic ice that dates as far back as 740,000 years, giving researchers a hint of temperatures and greenhouse-gas levels during the past eight ice ages. It took eight years, two attempts, and more than a bit of luck to extract the core. Initial tests suggest that our present interglacial period, like a similar one about 400,000 years ago, might last an exceptionally long time — another 16,000 years or so, without taking account of global warming.

In the hot seat

Climate researchers estimated that anthropogenic climate change has at least doubled the chances of a heatwave like the one that hit Europe in the summer of 2003. Although scientists have long thought that a warmer world will have more extreme weather, this result provided the most solid link so far between global warming and a single weather event. Such work could open the door for groups to win lawsuits against big emitters of greenhouse gases for damages caused by bad weather.

Table-top accelerators

High-quality electron beams for use in accelerators were produced by laser focusing. This eliminates the need for the massive magnets that have traditionally been used to focus such beams, and whittles down the instrumentation needed to make a particle accelerator from the size of a football stadium to the size of a lab. The ever-shrinking size of these devices will make them increasingly accessible to universities and individual research teams.

Home-made cure

A compound was mixed in the lab that could make for cheaper antimalarial drugs. Doctors currently recommend that high doses of artemisinin-based treatments be used in countries that have problems with resistance to drugs such as chloroquine. But artemisinin, which comes from a Chinese plant, is very expensive. A public-private partnership created a cheaper synthetic version, which is now in clinical trials in Britain.

I'd like to be able to get on with my experiments, which help people with dyslexia and Parkinson's disease, without being harassed by extremists.

John Stein
Neuroscientist, University of Oxford, UK.

When Stein this summer decided to join the handful of UK researchers who speak publicly about the benefits of animal research, he knew he would soon be the target of animal-rights activists. "I got a continuous stream of abusive e-mails," he sighs. One read: "What is the difference between a Nazi and a vivisector? Answer: nothing."

Stein and his colleague Tipu Aziz decided to speak up after activists scored two notable victories. In January, protests helped to force the University of Cambridge to abandon plans for a primate research centre. Six months later, the building contractor working on a new animal house at Oxford pulled out.

Both Stein and Aziz knew that taking a stand could be dangerous. In the past, at least, activists have gone beyond threats: supporters of animal research have been attacked with baseball bats and had letter bombs posted to their homes.

But the close of 2004 finds the pair

cautiously optimistic. Activists have now been portrayed as terrorists by some sections of the media, and their often militant approach has been exposed by undercover journalists. And in November, Oxford won a court injunction barring protestors from the immediate vicinity of the proposed site for its animal house. Stein and Aziz say that the threats they receive have now almost petered out.

They are also confident that the public is behind them — thanks in part to a strange experience Stein had at a Royal Institution event on nutrition and neuroscience this September. Stein and his brother Rick, a famous chef in Britain, co-hosted the event. They found themselves harassed by activists dressed in animal suits shouting that John was a "monkey torturer". But when police arrived, says John, they ended up having to protect the protestors from members of the public, rather than the Stein brothers from the protestors. Enraged, the public had turned on the picketers.

The recipe for a good 2005, say Stein and Aziz, would include more support from their university and fellow researchers in their efforts to explain why animal research is needed. Equally importantly, it would also involve the resumption of building work on Oxford's animal facility, which has been on hold since July. The university insists a new contractor will be found. But as *Nature* went to press, no builder had been named.

"For the United States to trump Europe and the rest of the world by announcing a successor to the Manhattan and Apollo projects: a bold initiative to decarbonize the energy system within five decades."

Stefan Rahmstorf
Climate modeller, Potsdam Institute for Climate Impact Research, Germany

If anyone has just cause to raise champagne glasses this New Year, it's the climate community. Researchers made significant progress in understanding the nature of past, present and future climate change (see Highlights, page 945). And after years of haggling, this autumn saw the Russian president, Vladimir Putin, ratify the Kyoto Protocol on Climate Change, finally establishing it as a legally binding international agreement. The Russian decision signalled much-needed political support for the push to reduce greenhouse-gas emissions — support

that seemed very far away this time last year. But, Rahmstorf and others note, the world will need more than Kyoto — preferably an effort with the United States fully involved — to get carbon dioxide emissions in check.

One scheme spurred on by the Kyoto Protocol looks set to make an impact in 2005: the idea of trading permits to emit carbon dioxide. About twice as many 'CO₂ equivalents' swapped hands in 2004 as in 2003. By January 2005, Europe's Emissions Trading Scheme will be in place, making carbon a government-regulated asset. Observers hope that giving emissions reductions a financial value will spur companies to cut down on their atmospheric garbage — even in countries that have not signed Kyoto, such as the United States. But no one yet knows if it will really reduce overall emissions.

Meanwhile, Hollywood's blockbuster *The Day After Tomorrow* gave the public a vivid snapshot of abrupt climate change, when it portrayed the entire Northern Hemisphere freezing solid in a matter of days. That may be a ridiculous exaggeration of what could happen in real life, but science continues to uncover evidence that dramatic temperature swings have occurred in the distant past, taking place over thousands or even hundreds of years. This year, for instance, researchers found hints of a warm Arctic climate in

the Cretaceous period some 70 million years ago. Data from samples taken during this year's Arctic Coring Expedition may soon tell us how the North Pole once turned into a mild Mediterranean basin.

"A big budget Hollywood movie epic that will make scientists the new idols of today's youth, causing a burst of interest in careers in science. Back off, rock stars, TV actors and athletes!"

Francis Collins
Director, National Human Genome Research Institute, Bethesda, Maryland

If *The Day After Tomorrow* wasn't good enough for Collins, there's plenty to look forward to — perhaps with mixed feelings. Among the films due out in 2005 is *Fantastic Four*, in which a group of astronauts gain superpowers after being exposed to cosmic radiation. Also set to hit the screens is the cult classic *The Hitchhiker's Guide to the Galaxy*, in which a spaceship with an improbability drive can do the seemingly impossible (as long as it knows how improbable it is). And in a new version of *War of the Worlds*, we will be treated to Tom Cruise's portrayal of life as a scientist. The gulf between celluloid and reality looks unlikely to be bridged any time soon, although even fantasy is sure to spark some interest in science — of a sort.

"A brightly coloured parrot that sits on my shoulder and every time I look at new data it screeches in my ear: 'But what does this really mean and is it important?'"

Brandon Wainwright,
human geneticist,
Institute for Molecular Bioscience, University of Queensland, Australia



"More uninterrupted thinking time."
Chris Rapley, director, British Antarctic Survey

"A bonfire of much of the idiotic new health and safety regulations that say I am supposed to put on a space suit before I can enter the animal house to study my chickens."

Steven Rose, neuroscientist, Open University, Milton Keynes, UK.

C. HUGHES/PANOS

"Very high on my wish list of discoveries for 2005 would be the development of a molecule that would elicit broadly reactive neutralizing antibodies against HIV."

Anthony Fauci

Director, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland

Those attending July's XV International AIDS Conference in Bangkok, Thailand, heard one refrain repeated over and over: we have the tools to treat the epidemic — but they're not reaching the majority of the 39 million people living with HIV. The epidemic is deeply entrenched in sub-Saharan Africa, home to more than 60% of people with HIV; and it threatens to take off in Russia and China, as well as India, which has the second highest number of HIV infections. But the antiretroviral medications that could treat patients are still too costly for most to afford. Faced with this daunting prospect, leaders are calling for preventative treatments that can stop the spread of AIDS once and for all.

But progress on an AIDS vaccine is slow. Although there are at least a dozen vaccines in clinical trials, researchers do not have high hopes that any will completely prevent people from contracting HIV. Many experts are convinced that what is needed to meet this goal is a vaccine that will stimulate antibodies that recognize and neutralize all forms of HIV. Structural biologists are now hard at work trying to identify these 'broadly neutralizing' antibodies.

*"A time controller.
This would allow —
at least subjectively
— the flow of time to
be increased,
decreased or stopped.
I believe some drugs
have this effect!"*

Arthur C. Clarke
Science-fiction author



"We are facing increasing risk of new emerging infections. I wish for constant vigilance, and for the resources to combat this threat with good science, surveillance and public policy based on science not politics."

Paul Tam

Acting pro-vice-chancellor, University of Hong Kong

Severe acute respiratory syndrome, or SARS, may now seem like a distant threat, with no new cases so far this winter. But there are other viruses to worry about. This year, bird flu led to the death by disease or slaughter of tens of millions of birds in countries across southeast Asia, and it killed at least 32 people in Thailand and Vietnam.

Evidence has emerged that the viral strain of greatest concern, H5N1, is present in pigs in China — an animal that could provide the perfect place for bird and human viruses to meet and mix, producing a lethal, highly transmissible version. Alarming, H5N1 seems to have passed from person to person in one case, when a Thai girl probably passed the disease to her mother. If the virus adapts to pass more easily between people, a deadly pandemic similar to those in the twentieth century is likely.

"The next pandemic is inevitable. In fact it's overdue," says David Ho, an infectious-disease expert at Rockefeller University in New York. And surveillance and healthcare systems in the developing countries may not be able to cope, he says.

Scandals

Missing plague

After reporting several vials of plague bacteria missing from his lab, and then admitting he might have accidentally destroyed them himself, US microbiologist Thomas Butler was sentenced to two years in prison for fraud this March. This is more lenient than the penalty sought by US prosecutors, who called for millions of dollars in fines and at least ten years in prison. But some researchers say it was unfair to make an example out of a 62-year-old, respected researcher with no terrorist ambitions.

For art's sake

A US university geneticist and an artist were accused of mail and wire fraud because of the way they allegedly obtained bacteria for art exhibitions. The investigation began when laboratory equipment, bacteria and books on biowarfare were found in the home of performance artist Steven Kurtz. The bacteria were found to be harmless, but both he and Robert Ferrell were accused of defrauding the supplier by using the organisms for non-research purposes outside the lab. Kurtz's case is set for a hearing on 11 January, while Ferrell's has been put on hold due to illness.

Autism paper 'flawed'

Medical journal *The Lancet* took the unusual step of distancing itself from one of its own papers and attacking its findings. In February, editors declared that Andrew Wakefield's 1998 paper linking the measles, mumps and rubella vaccine with autism was "flawed" owing to conflicts of interest and should not have been published. Wakefield said that there was no conflict. The paper caused many parents in Britain to decline the triple vaccine, and, as a result, measles incidence increased.

Cloning paper pulled

Fertility researcher Panayiotis Zavos had a peer-reviewed paper on human cloning pulled — because he publicized his work. Zavos created a fuss in the newspapers in September when he announced that he had created cloned embryos by mixing genetic material from dead people with cow eggs. The *Journal of Assisted Reproduction and Genetics* then pulled a paper on similar work, although Zavos claims it was a different study.

Locked out

A vetted Iranian physicist was banned from his workplace at the Stanford Linear Accelerator Center, a US Department of Energy lab in California. Colleagues told *Nature* that no explanation was offered for his expulsion. Shahram Rahatlou suspected his ban resulted from heightened security checks after 11 September 2001. Other Iranians said that it was now harder for them to work at, or even visit, government facilities. Rahatlou has since been offered a four-year position in Rome.

Odds & ends

Down with a bang

NASA's Genesis mission, which was designed to bring samples of the solar wind back to Earth, crashed into the Utah desert in September. The craft had such delicate detectors that mission designers had planned for Hollywood stunt pilots to swoop in and catch the capsule by its parachute, allowing for the softest possible landing. But the parachute didn't open, owing to an error in the design drawings that led to some crucial switches being installed upside down.

A whale of a time

An otherwise ordinary day in a busy Taiwanese street was interrupted in February by 60 tonnes of exploding sperm whale. The dead whale was being delivered by truck to a laboratory for an autopsy, when the carcass exploded after gas from decomposition built up inside. Luckily, only some of the internal organs fell into the street. The focus for the post mortem — the heart and lungs — was still intact.

Newton revealed

A 300,000-word interpretation of the biblical book of Revelation that Isaac Newton wrote in the late seventeenth century was published online in August. The eye-opening text, peppered with references to dragons and reflections on distrust of the Catholic faith, revealed Newton's intense interest in spiritual matters. More than half of Newton's works seem to have been predominantly about religion rather than science.

From the archive

A theoretical physicist aired his life history, including stories of growing up as a Polish Jew in occupied France during the Second World War, in an unusual medium this year. He put it all on the arXiv physics preprint server, which more usually hosts original research. But the archive won't put everything up online. Researchers who feel they have been unfairly excluded from the server banded together this year to form their own 'archive freedom' website, to protest at the site's selection criteria.

Beagle wrangling

No one really knows what happened to the ill-fated Beagle 2 lander when it went missing on its descent to Mars this time last year. But the bigger question may be who will pay for the failed attempt. The European Space Agency (ESA) 'lent' mission leaders in the United Kingdom €16 million (US\$21 million) for the project, and this autumn some ESA-funded researchers were beginning to grumble about whether, and how, the space agency would ever get it back. Britain may or may not want to repay in kind with goods and services, but space scientists were quoted as saying a cheque would be best.

"To find an organism in ocean sampling that would help to eliminate the world's dependency on carbon-based fuels."

Craig Venter, head, J. Craig Venter Science Foundation, Rockville, Maryland

"Five minutes with Charles Darwin. Or, failing that, a modern genomics laboratory for the Charles Darwin Research Station in the Galapagos."

Hunt Willard
Geneticist, Duke University, Durham, North Carolina.

The Galapagos Islands are a living laboratory for studying evolution, and naturalists have been drawn there for more than a century. The islands once helped spark revolutionary theories about evolution, but experts say that the archipelago's Charles Darwin Research Station on Santa Cruz now desperately needs better equipment to keep pace with twenty-first-century genetics — and to unravel some remaining mysteries of evolution.

Biologists tend to assume that evolutionary changes arise from rare genetic mutations becoming fixed in a species population thanks to environmental pressures. But, says Willard, changes on the Galapagos Islands seem to be happening too quickly for this to be the only mechanism at work. More dramatic, large-scale genomic changes may be occurring as a result of coupling between different subspecies, he says. But until a 'Galapagos revisited' project does thorough genomic studies, we won't know what's really going on at the DNA level, he says.

Such genetic studies have already proved useful for some of the lonelier species on the islands. An analysis of tortoises helped to find the best potential subspecies match for Lonesome George — a giant tortoise thought to be the last of his line on the island of Pinta. George has not yet been introduced to any of these chosen females, but scientists are hoping that he won't ignore them in the same way that he has rejected the other potential mates now in his enclosure.



"A spell check for English (Euro-speak) to add to my computer languages of English (UK) and English (US). And a dictionary to go with it, so I can work out what the Euro-words actually mean."

Anon

Coffee-breaks at European science conferences this year were alive with

complaints about the increasingly difficult application process for European 'Framework' grants. Scientists say that the forms, riddled with neologisms such as 'sideground', are becoming increasingly impossible to read, let alone to fill out.

There actually is a glossary for the Framework programme (see <http://fp6.cordis.lu/fp6/glossary.cfm>), but sadly neither it nor the *Oxford English Dictionary* includes the word 'sideground'. If you dig through the European Commission's Intellectual Property Rights Helpdesk website, you'll find that the word means "information and rights acquired in parallel with a project".

Scientists say that the increasingly complex application forms seem to want them to prove that they will help cure Europe's economic and social ills, while doing a bit of science on the side. Writing the research project is the easy part, they say; trying to work out how to handle the political add-ons is a full-time job.

The details of the Seventh Framework Programme, to begin in 2006, will be hammered out during 2005. But it's a good bet that any wish for simplicity in the new application forms won't be granted. Researchers instead set their hopes on the creation of the planned European Research Council, which should be distanced from politics in Brussels.

Add your thoughts to *Nature's* wish list at
www.nature.com/news/specials/wishlist

"A firm commitment by the European Commission to earmark enough money for the European Research Council, no matter whether the research budget will be doubled or not."

Erwin Neher, biophysicist, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Progress requires scientific thinking at all levels

This underpins the core values of a moderate society and must not be limited to an élite.

Sir—As a medical student and a Pakistani, I was encouraged to read the Commentary by Atta-ur-Rahman and Anwar Nasim, “Time for ‘enlightened moderation’” (*Nature* **432**, 273; 2004). Students would be the first to benefit from the development of a scientific culture across the Islamic world. On the other hand, increasing spending on higher scientific education (by funding PhDs and establishing research centres, as the authors suggest) would not be an effective way to develop a moderate, mature and tolerant Islamic society.

Instead of offering narrow solutions that largely benefit those within the scientific community, we would do better to encourage

scientific thinking across all segments of society, so that intellectual rigour and open-mindedness become core social values.

Scientific thinking is characterized by scepticism, objectivity, an appreciation of uncertainty and the flexibility to alter one's beliefs in the face of conclusive evidence. Within the scientific community, these values stimulate open debate and ensure rigorous analysis of data and hypotheses.

However, if public interest in science is underdeveloped and the values underlying scientific thinking remain alien to large portions of the populace, scientists will be unable to effect any real change in social attitudes.

To develop a moderate, enlightened society, the values that underpin scientific thinking should be cultivated among secondary-school students, by improving the quality and availability of scientific education for everyone.

I do not doubt the importance of an active, productive scientific community for the intellectual and economic well-being of a nation. But I do not believe that expending resources to advance the scientific education of a few will bring enlightenment to a whole society.

Sheheryar Kabraji

Wolfson College, Barton Road,
Cambridge CB3 9BB, UK

Meyer paper: don't hang the Soc. Wash. out to dry

Sir—As a systematist I was dismayed to read Vladimir Svetlov's comments in Correspondence (*Nature* **431**, 897; 2004). Svetlov addresses the unfortunate publication of S. C. Meyer's ‘intelligent design’ (ID) paper (*Proc. Biol. Soc. Wash.* **117**, 213–239; 2004). However, instead of directing criticism at that paper, he makes scathing claims about low-impact journals in general and the *Proceedings* in particular.

Svetlov describes the *Proceedings* as a journal that “enjoyed much-deserved obscurity”. This characterization is not accurate. A cursory review of authorship in the *Proceedings* throughout its 122-year history reveals a list of everyone who's anyone among systematic biologists, including scores of notable past and current scientists from the Smithsonian Institution.

Journals that primarily publish taxonomic descriptions (such as the *Proceedings*) generally have low impact factors, but the relevance of such papers is often long-lasting relative to those in high-impact journals, as they are cited across decades and centuries rather than over a period of a few years (a search of *Nature*'s website shows that *Proceedings* articles have been cited in at least three Letters to *Nature* since 2002). Such papers are hardly “inconsequential”.

Svetlov says: “The editors and reviewers of many low-impact journals cannot provide the quality reviewing process one gets with *Nature*, *Science*, *Cell* and a few (very few indeed) other established magazines.” My own experience is that a journal's impact factor does not reflect the quality of the review process per se. Submissions to low-impact periodicals are

often reviewed by sticklers who examine mundane conclusions with the same caution that a reviewer for *Nature* would use in evaluating more grandiose scientific claims. Indeed, the same experts commonly evaluate papers in both high-impact journals such as *Nature* and low-impact specialist journals. This is why Svetlov should perhaps not be “surprised it took so long” for a paper such as Meyer's to appear in the peer-reviewed literature.

Given the *Proceedings*' taxonomic focus, Meyer's ID paper is clearly out of place. Its publication represents a lapse of the journal's usual editorial policies, and has been swiftly repudiated (www.biosocwash.org). However, although the publication of Meyer's paper is lamentable, it need not be used to trivialize the *Proceedings*' long, respectable and ongoing tradition of cataloguing global biodiversity.

Kristofer M. Helgen

School of Earth and Environmental Sciences,
University of Adelaide, Adelaide, SA 5005,
Australia, and South Australian Museum,
North Terrace, Adelaide, SA 5000, Australia

Meyer publication worse than just bad science

Sir—Vladimir Svetlov makes some interesting points, in Correspondence, about the proliferation of peer-reviewed journals and the publication of flawed papers (*Nature* **431**, 897; 2004). But he does not take the recent publication of an ‘intelligent design’ (ID) paper (*Nature* **431**, 114; 2004) seriously enough.

We agree that the paper presented no new arguments and appeared in a relatively obscure journal. For such reasons it is unlikely to influence scientists. However,

this does little to diminish its usefulness to ID proponents, who wish to influence public rather than scientific opinion.

The point is that, before it was withdrawn, this ‘peer-reviewed publication’ could be used by ID supporters in the United States to lend apparent legitimacy to their efforts to convince legislators and state and local education boards that ID is science and should be taught alongside darwinian evolution in public schools.

Such efforts are, alarmingly, bearing fruit in many US states — including Ohio (see www.ncseweb.org/resources/news/2004/OH/832_critical_analysis_of_evolution_3_10_2004.asp), where Svetlov himself is currently based.

Day B. Ligon, Matthew B. Lovern

Department of Zoology, Oklahoma State University,
430 Life Sciences West, Stillwater,
Oklahoma 74074, USA

Let's see what happens if I press this button...

Sir—I simply cannot free my mind of the marvellous dreamscape inspired by the Correspondence letter “US rules on tech transfer to foreign nationals” from Peter Lichtenbaum of the US Department of Commerce (*Nature* **432**, 15; 2004).

I, the humble foreign-national visitor, enter the generous host laboratory and start twiddling the knobs on their latest \$1-million gizmo, perhaps with disastrous results for its integrity.

But the resident expert technician cannot approach me with advice because that's ‘training’, and requires a licence!

Robin Thompson

Five Halls, Little Bookham Common,
Surrey KT23 3HY, UK

Beyond the culture shock

Culture is a key ingredient in the development of human societies.

Not by Genes Alone: How Culture Transformed Human Evolution

by Peter J. Richerson & Robert Boyd
Chicago University Press: 2004. 344 pp.
\$30, £21

Robin Dunbar

I continue to be surprised by the number of educated people (many of them biologists) who think that offering explanations for human behaviour in terms of culture somehow disproves the suggestion that human behaviour can be explained in darwinian evolutionary terms. Fortunately, we now have a book to which they may be directed for enlightenment.

The authors made their name 20 years ago with the publication of *Culture and the Evolutionary Process* (Chicago University Press, 1985). Alas, it was one of those classic books that everyone quotes with approbation but, because it contains some hard mathematics, few have had the stomach to read in any detail. If you fall into that category, this anniversary volume is what you have been waiting for: there is not a single equation to disturb the tranquility of

your reading, not even a graph. Richerson and Boyd have gone to great pains to present the core of their original theory of cultural evolution and elaborate its implications for the wider study of human behaviour in a form that even those with no mathematical background will find both engaging and enlightening.

Richerson and Boyd's position, in a nutshell, is that culture is as good a candidate for darwinian treatment as behaviour and morphology. Genetic processes and individual learning provide mechanisms whereby successful phenotypes can be passed on from one generation to the next. In humans, and possibly in other species, including apes and cetaceans, culture provides another such mechanism for behavioural phenotypes. It is true that, in humans, cultural inheritance has been raised to an art form. But all this means is that the evolutionary dynamics that arise from both its intrinsically different modes of inheritance (sideways as well as vertically) and its complex interactions with more conventional genetic mechanisms make human behaviour a more challenging — and therefore more interesting — field

for evolutionary biologists to explore. "The existence of culture," comment Richerson and Boyd, "is a deep evolutionary mystery on a par with the origin of life itself."

Drawing on new ideas about multilevel selection, evolutionary psychology and what has become known in the trade as 'strong reciprocity' (the bestowing of rewards and punishments even where there is no direct personal gain for this behaviour), Richerson and Boyd build a case for a special role for cultural processes in human evolution. Their position is founded on an evolved psychology that predisposes humans to use imitation as a quick and inelegant way of cutting through the costs of obtaining information about the world first-hand. This is useful whenever environments change slowly but the information available about the change is poor or costly to obtain. Culture may be maladaptive on occasion, but that is not an evolutionary issue — something that non-evolutionists invariably fail to understand. Evolutionary explanations are statistical by their nature, and depend on the balance of the costs and benefits.

All this is a build-up to the big story of



Sharing a drink: altruism and group identity bind together societies such as the Nenets herders in Siberia.

BRYAN & CHERRY ALEXANDER PHOTOGRAPHY

the book, which is how to explain hierarchical group structures in traditional human societies, and the forms of indiscriminate altruism that often accompany them. (Richerson and Boyd seem to want to use the word 'always' here but, in a rare lapse of concentration, I think they seriously overstep the evidence.)

Conventional evolutionary explanations, such as kin selection and reciprocal altruism, work well for small-scale societies, like those found among monkeys and apes. But they do not work quite as well on the larger scale of human tribal societies, where kinship and interaction frequencies are too dilute to sustain indiscriminate altruism. When individuals' ecological and reproductive success depend on the success with which they can cooperatively solve the problems of survival, something is needed to prevent the age-old tension between self- and group-level interests tearing the fabric of groups apart. Something is needed to enforce commitment to the group, and that something, Richerson and Boyd argue, is culture, creating a sufficiently strong sense of group identity and conformity to enable groups to do their job for the individual.

In many ways, this book is really addressed to social scientists, who, as Richerson and Boyd remark, learned their evolutionary biology from the mischievous writings of Steven Jay Gould, but failed to realize that his seductive polemics have little empirical support. Functional explanations for natural phenomena are, they remind us, often difficult to see without a great deal of hard empirical work. Gouldian arguments of 'spandrels as non-adaptation' are, they remind us, usually at least as much 'just so stories' as adaptationist explanations — if not more so, as they are rarely backed by anything resembling empirical evidence. And it is evidence, they insist, that is the core to good evolutionary biology. No amount of armchair theorizing can substitute for hard-earned natural-history knowledge, as the seminal examples of Aristotle (whose biology, incidentally, was outstanding, however bad his physics might have been) and Darwin remind us.

In many ways, this book is a plea to the social sciences to take seriously what evolutionary biology has to offer — to see it not as a threat intent on cannibalizing the social-science niche, but rather as a source of useful techniques and conceptual devices that can be applied within a conventional social-sciences framework, asking conventional social-science questions. It is a book full of good sense and the kinds of intellectual rigour and clarity of writing that we have come to expect from the Boyd-Richerson stable. ■

Robin Dunbar is in the School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK.



DRAC RHÔNE-ALPES, SERVICE REGIONAL DE L'ARCHEOLOGIE

Quick on the draw? The age of paintings of horses at Chauvet cave in France is controversial.

Charting the past

The Seventy Great Inventions of the Ancient World

edited by Brian M. Fagan
Thames & Hudson: 2004. 304 pp.
 £24.95, \$40

Paul G. Bahn

The 'Seventy Great' book franchise, which first appeared a few years ago, has until now seemed rather lightweight, focusing as it has on 'wonders' and 'mysteries'. The volumes were uniformly attractive and often well written, but they inevitably reminded one of those pointless 'top 10' and '100 best' shows that help to pad out the British TV schedules. It is likely that no two archaeologists would ever reach a consensus on which topics to include in such compilations.

But this new addition to the list is a distinct improvement: as a survey of ancient technology, it is far less random and certainly covers the basics. Edited by Brian Fagan, an eminent popularizer of archaeology, it presents almost all the major inventions of humankind, from its origins up to an undefined cut-off point. The 70 topics, we are told, were chosen from a potential list of hundreds, but the basis on which the selections were made, or who made them, is not revealed. Nevertheless, I think archaeologists would agree that few important topics have been omitted, and the team of 42 authors certainly includes a number of outstanding specialists.

Overall, the volume is very up-to-date and includes recent finds such as Germany's Nebra sky disc. Some of the essays are extremely good, notably those on pottery, watercraft and fermented beverages. Others are weaker, but this is an inevitable problem in such works — nobody can really be an expert on every period, and it shows.

In particular, some essays, such as those on clothing or jewellery, display a distinct

lack of knowledge of the Stone Age evidence, despite the earliest periods being the most crucial to the book as a whole, as this is when these and many other inventions actually arose. For example, astonishingly, the essay on clothing makes no mention of Italy's Neolithic Iceman, who has made a huge contribution to our knowledge of early everyday wear, as opposed to burial garments; and the great Neolithic wooden well of Kückhoven, Germany, is absent from the essay on water supplies. The piece on boomerangs ignores not only those found in Tutankhamen's tomb, but also the great ivory specimen from Oblazowa, Poland, dating to at least 18,000 years ago; it may have been a ceremonial object, but nevertheless, trials with a facsimile proved it to be an efficient and aerodynamic weapon.

Among the topics that are missing completely, but which could well be argued to have been far more important to human life than board games, ball games, or codes and cyphers (all of which are included), one might mention levers and pulleys, ladders and scaffolds, steps and staircases, cordage and knots — and language. Other topics that are mentioned only briefly, such as mirrors and dentistry, should surely have been accorded much greater significance.

The book is also strangely inconsistent in its attitude to evidence. Some highly controversial items are accepted without a murmur, such as early dates for France's Chauvet cave, a late date for people reaching the Americas, and the Neanderthal 'flute' of Divje Babe, Slovenia, which is almost certainly natural, made by carnivores. Yet the early proto-figurine of Berekhat Ram, Israel, is treated with great scepticism, despite having been the subject of two microscopic analyses.

But overall I can recommend this book as an entertaining pot-pourri. It presents a great deal of fascinating information in a highly readable and well illustrated way. ■

Paul G. Bahn is a freelance archaeologist at 428 Anlaby Road, Hull HU3 6QP, UK.

Science in culture

Kepler's seasonal gift

What lies behind the hexagonal shape of the snowflake?

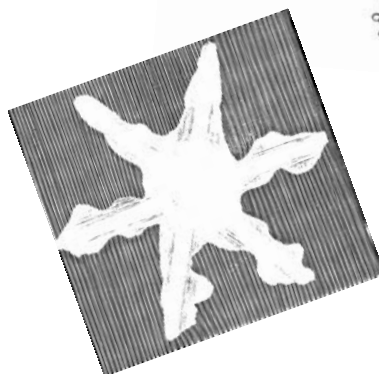
Martin Kemp

In 1611 Johannes Kepler, the German cosmologist with a passion for geometry, offered his patron, John Matthew Wacker, a highly seasonal "New Year's gift". Unlike most such gifts, it not only outlasted the festive season but also acquired enduring renown. Kepler's present was his small book, *On the Six-Cornered Snowflake*. It was a sprightly and speculative expression of delight and perplexity in the face of the marvellous regularity of snow crystals.

Kepler provided no illustration, as his patron well knew what a hexagon looked like. But Kepler was unaware of the astonishing variety within the crystalline structures, which was to be revealed by the microscope and illustrated in Robert Hooke's *Micrographia* in 1665, as shown here.

The account by Kepler of what triggered his thoughts on the snowflake gives a nice idea of his book's tone: "Specks of snow fell here and there on my coat, all with six corners and feathered radii. 'Pon my word, here was something smaller than any drop, yet with a pattern; here was the ideal New Year's gift for the devotee of Nothing'." The jest that his patron was a "devotee of Nothing" may be a philosophical play on the fact that the fundamental unit of pure geometry, the point, is 'nothing', as it has no physical dimension.

Kepler tries to identify the cause of the snowflake's six-cornered configuration. But before doing so, he notes the ubiquity of polygonal and polyhedral constructions in nature, from the cosmos to the world in miniature. He is delighted to observe how the polyhedral architecture of the snowflake has analogies across all scales of natural design. He had himself argued that the



geometry of the five 'platonic solids', or regular polyhedra, provided a kind of invisible armature for the orbits of the planets (see *Nature* **393**, 123; 1998). On a smaller scale, the rule of geometry is manifest in the world of plants, Kepler points out: "in a flower the authentic flag of this faculty is flown, the pentagon". The classic case is the bee's honeycomb: "The architecture is such that any cell shares not only six walls with the six cells in the same row, but also three plane surfaces on the base with three other cells from the contrary row."

Such geometrical packing, which Kepler compares to the seeds in a pomegranate, is attributed to a physical process similar to that observed when pellets are systematically compressed in a round vessel. But the regularity of the physical actions must themselves be caused by something, and Kepler concludes that God prescribed to the bee "these canons of its architecture" that resulted in the geometrical constructions.

But what lies behind the form of the snowflake,

when no animate architect is involved? Observation and delight are one thing; explanation, beyond the adducing of divine agency, is very much another. Kepler's review of a range of physical explanations for the forms of the snowflake proves frustratingly inconclusive. He is convinced that "the cause of the six-sided shape of a snowflake is none other than that of the ordered shapes of plants and numerical constants". In the absence of any obvious agent, he adduces the presence of a "formative faculty" — the *facultas formatrix* — which God has insinuated into matter in terms of "world-building figures". But, in the final analysis, Kepler senses that something is missing in his explanation, and he concludes by announcing that he has "knocked at the door of chemistry".

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK, and co-director of Wallace Kemp Artakt. His book *Leonardo* was recently published by Oxford University Press.

Beginning again

Big Bang: The Most Important Scientific Discovery of All Time and Why You Need to Know About It

by Simon Singh

Fourth Estate: 2004. 544 pp. £20, \$27.95

Peter Coles

When the British astrophysicist Fred Hoyle coined the phrase 'Big Bang' to describe the rival to his beloved 'steady state' theory of the Universe, he meant it to be disparaging. It was bad enough for Hoyle that his pet theory turned out to disagree with astronomical observations, but it must have been especially galling that his cosmological adversaries embraced his derisive name. The tag has since spread into the wider cultural domain — nowadays even politicians have heard of the Big Bang.

But what is the Big Bang? In a nutshell, it is the idea that our Universe — space, time and all its matter content — was born in a primordial fireball, from which the whole caboodle has been expanding and cooling ever since. Pioneering theorists such as Aleksander Friedmann and Georges Lemaître derived mathematical solutions of Einstein's field equations that could be used to describe the evolution of a Big Bang Universe. These models involve a creation event, in which space-time and matter-energy sprang into existence to form our Universe. We are still in the dark about how this happened, but we think it took place about 14 billion years ago.

Edwin Hubble's discovery of the recession of distant galaxies gave support to the idea that the Universe was expanding, but the notion that it might be evolving from a hot beginning was rejected by many theorists, including Hoyle. He favoured a model

in which the origin of matter was not a single event but a continuous process in which atoms were created to fill in the gaps created by cosmic expansion. The battle between these competing views of creation raged until the accidental discovery in 1965 of the cosmic microwave background radiation, which marked the beginning of the end for the steady-state theory.

This conflict between the two theories plays a central role in Simon Singh's book *Big Bang*. His previous books, *Fermat's Last Theorem* and *The Code Book*, succeeded admirably in bringing difficult mathematical subjects to a popular readership, using a combination of accessible prose, a liberal sprinkling of jokes and a strong flavouring of biographical anecdotes. The recipe for his new book is similar.

In *Big Bang*, Singh uses the historical development of modern cosmological theory as a case study for how scientific theories

are conceived, and how they win or lose acceptance. He rightly points out that science rarely proceeds in an objective, linear fashion. Correct theories are often favoured for the wrong reasons; observations and experiments are frequently misinterpreted; and sometimes force of personality holds sway over analytic reason. Because cosmology has such ambitious goals — to find a coherent explanation for the entire system of things and how it has evolved — these peculiarities are often exaggerated. In particular, cosmology has more than its fair share of eccentric characters, providing ample illustration of the role of personal creativity in scientific progress.

This very well written book conveys the ideas underpinning cosmological theory with great clarity. Taking nothing for granted of his readership, Singh delves into the background of every key scientific idea he discusses. This involves going into the history of astronomical observation, as well as explaining in non-technical language the principles of basic nuclear physics and relativity. The numerous snippets of biographical information are illuminating as well as amusing, and the narrative is driven along by the author's own engaging personality.

As a fan of Singh's previous books, I have to admit that, although this one has many strengths, I found it very disappointing. For one thing, there isn't anything in this book that could be described as new. The book follows a roughly historical thread from pre-classical mythology to the middle of the twentieth century. This is a well-worn path for popular cosmology, and the whole thing is rather formulaic. Each chapter I read gave me the impression that I had read most of it somewhere before. It certainly lacks the ground-breaking character of *Fermat's Last Theorem*.

The past ten years in cosmology have witnessed a revolution in observation that has, among many other things, convinced us of the existence of dark energy in the Universe. Theory has also changed radically over this period, largely through the introduction of ideas from high-energy physics, such as superstring theory. Indeed, some contemporary Big Bang models bear a remarkable resemblance to the steady-state universe, involving the continuous creation not of mere atoms, but of entire universes.

Frustratingly, virtually all the exciting recent developments are missing from this book, which leaves off just when things started to get interesting, with the COBE satellite in 1992. Readers who want to know what is going on now in this field should definitely look elsewhere. The processes of cosmic discovery and controversy are ongoing, not just relics of the past. ■

Peter Coles is in the School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG9 2HL, UK.



G. LIGER-BELAIR

Star quality: bubbles at the surface of champagne arrange into vortices that look like galaxies.

..... Fizzical attraction

Uncorked: The Science of Champagne

by Gérard Liger-Belair

Princeton University Press: 2004. 152 pp.

\$19.95, £12.95

Richard N. Zare

This book presents the birth, life and death of a champagne bubble with such gusto, good humour and clarity that you will devour its delicious contents in one gulp. Whereas good champagne is to be sipped, this book is not. You will never experience the sensual elegance of champagne in quite the same way again once you have read this entertaining account of its history and 'fizzics'.

The author is associate professor of physical sciences at the University of Reims Champagne-Ardenne and consultant for the research department of champagne house Moët & Chandon. He brings to this topic not only great expertise, but also a fine sense of aesthetics, as he shares his striking photographs of how a bubble forms, rises and bursts. These phenomena are often much more complex than we imagine, but Liger-Belair explains bubble science without resorting to a single equation. He succeeds in clarifying for the non-expert such abstruse topics as fluid dynamics, nucleation phenomena, fermentation and the competition between buoyancy and drag. *Uncorked* provides just enough science to be authoritative and instructive, and carries the reader enthralled from one chapter to another, like a good detective story.

Liger-Belair has made several important original contributions to the science of champagne. For example, he offers photographic proof that bubbles form more on impurities — often fibres that adhere to the walls of the champagne container (a fluted glass, say) — than on microscratches and

irregularities in the container walls. These hollow fibres, which act as tiny bubble guns, are deposited on the walls when the champagne flute is dried with a paper towel or cloth. Indeed, without these defects, champagne would not fizz at all. But bursting bubbles are not the only way champagne loses its sparkle. The dissolved carbon dioxide escapes primarily from the liquid surface of the champagne, rather than from the popping of its many tiny bubbles.

Many mysteries are revealed. Why do bubbles rise more rapidly in soda water than in champagne, and more rapidly in champagne than in beer? All three liquids are mostly water and have approximately the same viscosity. For a solution to this riddle, read *Uncorked*.

For my own taste, I would have preferred the discussion to have stuck to champagne. Instead, the author ends with an afterword on the future of champagne wines, emphasizing the connection of the human production of carbon dioxide to global warming. He concludes with some speculation: It is "not completely unrealistic to think that we eventually may witness the emergence of fine sparkling wines in Great Britain". Climate change deserves its own book, a much larger one of a more serious nature; the emphasis on it seems out of place in this vignette, although Liger-Belair's concern about our effect on the environment is shared by almost all scientists.

Many scientists feel that physical principles are part of our daily lives, not just something to be studied in the laboratory. This book reinforces that belief. It would make an excellent Christmas gift to anyone who wants to learn how a scientist views the world — or to anyone who has ever wondered about the mysterious fluid that has launched so many ships and toasted so many celebrations. ■

Richard N. Zare is in the Department of Chemistry, Stanford University, Stanford, California 94305, USA.

Reality is wilder

How a terrifying interview led to fun doing physics.

Gregory Benford

Job interviews can be harrowing, but few ultimately change one's career directions. One afternoon in 1967, a personnel clerk at the Lawrence Radiation Laboratory, California, ushered me into a large, messy office, saying only that 'someone' wanted to ask me a few questions.

I was being interviewed for what was then (and still is) the best-paying postdoctoral position in physics. About to become a freshly minted PhD from the University of California, San Diego, I wondered if this was about security clearance. Instead, a distracted Edward Teller sat behind an untidy desk that was piled high with physics journals. Nobody had told me that Teller insisted on taking the measure of every postdoctoral candidate. "We didn't want you to be nervous," someone said later. It worked — I was merely terrified.

He was the most daunting job interviewer imaginable, a famous physicist and director of the laboratory, looming large in a central mythology of modern science — the A-bomb. In that next hour, no one disturbed us as Teller quizzed me in detail about my thesis in solid-state plasmas. Attentively, he turned every facet over and over, spying undiscovered nuances, overlooked difficulties, a calculation a bit askew. I struggled to keep up, incoming questions splattering like bugs on my conceptual windshield. By the time you see the problem, it's too late.

My thesis explored how strong magnetic fields could bind extra electrons to hydrogen-like impurities in indium antimonide — an effect later verified. Teller saw many implications, and soon stood beside me at the blackboard, dashing out equations. He was brilliant, leaping ahead of my jittery explanations to see connections that I had only vaguely sensed. His mind darted as swiftly as any I had ever encountered. Within minutes, I was sweating. That one hour lasted days.

To my vast surprise, I apparently passed inspection. At the end, there was a long pause and he gestured for me to sit in a wooden chair beside the blackboard. I sank into it with relief, showering chalk from my hands onto his desk. He then announced, "The most important question of all," and, leaning closer with a sudden scowl, then asked, "Will you be willing to work on whatever comes up?"

Unbidden, images from Stanley Kubrick's film *Dr. Strangelove* leaped to mind. He had already asked if I would be willing to work on



Edward Teller (above) cast an imposing presence over modern physics, intimidating the author (below).

weapons, but this was even more open-ended. However, Teller had impressed me as a deep, reflective man. I had grown up in the shadow of the cold war. My father was a career army officer and our family had lived in occupied Japan and Germany. These were advanced nations, yes, but the greatest can blunder the most. It seemed to me that the sheer impossibility of using nuclear weapons was the best, indeed the only, way to avoid strategic conventional war, whose aftermath I had seen in shattered Tokyo and Berlin.

So I agreed. There was no Dr. Strangelove. Teller just wanted to do physics and have fun. Within a few weeks, I had a job offer — for whatever came up. In the next four years at Livermore, I had a chance to work with Teller and the other wild talents, such as Lowell Wood, who relished the collision of colourful imagination and careful analysis.

I got involved with the study of tachyons, the theoretically possible particles that can travel faster than light. Not the sort of thing one imagines a 'weapons lab' doing, but Teller allowed the theorists a wide range. When the tachyon idea popped up in the physics journals, Teller thought they were highly unlikely, and I agreed — but worked on them anyway out of sheer speculative interest. Moving faster than light implies particles can travel backwards in time.

Teller invoked a different argument against tachyons. Echoing Enrico Fermi's famous question, "Where are they?" that asked why

aliens, if they are plentiful in the galaxy, haven't visited us by now, Teller noted that if tachyons existed, "Why haven't they been sent? Where are our messages from the future?"

My answer was that nobody had built a tachyon receiver yet. Neat, but perhaps too neat. Surely nature would not disguise such a profound effect. With Bill Newcomb and David Book, I published the paper 'The Tachyonic Antitelephone' (*Phys. Rev. D* 2, 263–265; 1971). We destroyed the existing arguments, which had avoided time-travel paradoxes by reinterpreting tachyonic trajectories moving backward in time as their antiparticles moving forward in time. Without using a single equation, we showed that imposing a signal on the tachyons — sending a message — defeated the reinterpretation, so causality problems remained. Time plays a strange role in physics. Paradoxes are hard to rule out. I later wrote a novel about this, *Timescape*, which has been in print a quarter of a century, suggesting that fundamental puzzles remain.

This pattern, speculation leading to detailed theory, is one I have often encountered. I even wrote novels about it, a hobby that has at times taken my career in odd directions. Ideas blossom from imagination, but they must be tested against reality; that, as I learned from Teller, is the essence of science. The received wisdom of science is quite prissy, speaking of how anomalies in data lead theorists to explore models, which are then checked by dutiful experimenters, and so on. Reality is wilder than that. ■

Gregory Benford is in the Department of Physics, University of California, Irvine, California 92697-4575, USA.

LAWRENCE LIVERMORE NATIONAL LABORATORY

J. BENFORD

Shock fronts in Hawaii

Alan P. Boss

The most vexing question in meteoritics is on the verge of being answered — what process led to the small droplets of primordial dust that are found throughout the most primitive meteorites?

The meteorites and comets that were formed in the solar nebula, the planet-forming disk of gas and dust that became our Solar System, have been largely unaltered over time, and their structure provides clues about the physical and chemical processes that occurred in the nebula. One group of early meteorites, chondritic meteorites, are composed largely of chondrules — millimetre-sized spheroids with well-defined edges resembling “drops of fiery rain”¹.

Understanding how the chondrule precursor grains were heated to melting has long been the top priority of many meteoriticists, and they met last month in Kauai, Hawaii*, to tackle the problem. According to results presented at the meeting, some 4.6 billion years ago our inner Solar System must have seen episodic fiery rain showers of molten chondrules that lasted for a million years or more. It could be that these persistent showers were produced by shock fronts originating from spiralling arms in the solar nebula.

The chondrules' peculiar textures provide information on the rate at which they were heated to close to their melting temperatures. These textures show that flash heating² must have been involved: the precursor dust aggregates were heated to somewhere between 1,800 K and 2,200 K and cooled back down in a matter of minutes (R. Hewins, Rutgers Univ.). The problem is to find the origin of such a violent event. An added difficulty is that the flash heating must have been repetitive, as some chondrules seem to have melted many times.

The search for a reliable flash-heating source for chondrules has been a long one, with mechanisms proposed ranging from impact melts produced by collisions between the growing rocky bodies of the inner Solar System, to lightning strikes within the solar nebula. The consensus, however, is that shock fronts within the nebula were responsible^{3,4}: the chondrule precursors

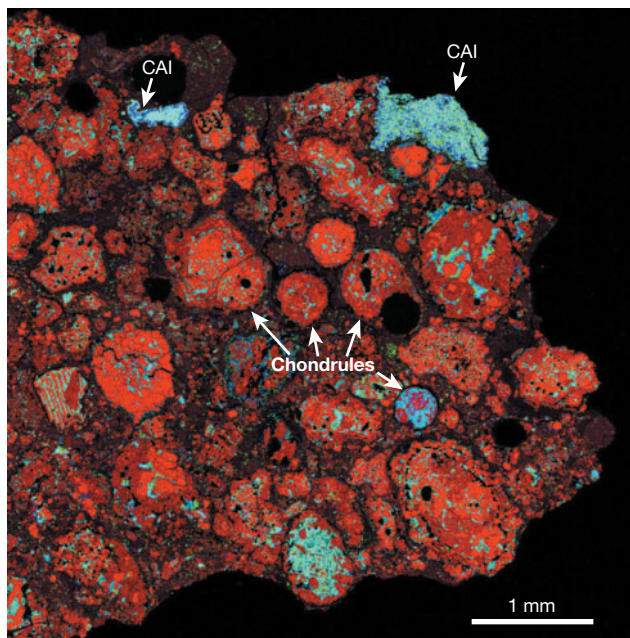


Figure 1 A chondritic meteorite containing both chondrules and calcium-and-aluminium-rich inclusions (CAIs). This X-ray cross-section of meteorite PCA 91082 shows the elemental abundances in the meteorite: red is magnesium, green is calcium and blue is aluminium. The chondrules seem to have been heated to melting by shock fronts in the solar nebula, and then mixed together with CAIs to form the chondritic meteorites. (Courtesy of A. N. Krot, Univ. Hawaii.)

were heated to near melting temperatures by frictional gas drag as they struck shock fronts travelling at speeds of 6–9 km s⁻¹, as well as by radiation from the hot shocked gas and other chondrules (S. Desch, Arizona State Univ.). But what produced these long-lasting shock fronts? Without a source for their regeneration, they would be expected to disappear quickly.

Whatever the process was, it seems to have continued for a million years or more. This conclusion follows from studies of long- and short-lived radioactive isotopes, which provide a precise chronology for the early Solar System^{5,6}. Such studies show that some chondrules formed at the same time as the earliest solids, the so-called calcium-and-aluminium-rich inclusions (CAIs; Fig. 1), whose ages define the 4.567-billion-year age of the Solar System⁷. Other chondrules formed as much as a million years or so after the CAIs (Y. Amelin, Univ. Toronto). Some CAIs also seem to have been flash heated after they formed.

One possible source of shock fronts in the

nebula is planetesimals of 1,000-km diameter moving in highly eccentric orbits (L. Hood, Univ. Arizona). However, bodies of that size would not be expected to have such eccentric orbits unless their orbits had been affected by Jupiter, and the standard picture is that Jupiter took about 4 million years to form⁸ — too long to account for the formation of the earliest chondrules. This model is also incompatible with the general belief that chondrules began to form, along with the earliest solids, before the formation of planetesimals.

The strongest candidate for the source of shock fronts is spiral arms in the solar nebula⁹. Different theories of Jupiter formation point to the same conclusion — that the disk-shaped nebula must have been slightly gravitationally unstable, with the result that spiral arms in the outer regions would have formed spontaneously and collided to produce Jupiter-mass clumps. Regardless of whether or not those clumps survived and formed Jupiters, they would have driven one-armed shock fronts down to

the very centre of the nebula. Because these shock fronts were tied to clumps orbiting at Jupiter's distance and beyond, the solids moving with the gas at asteroidal distances would have encountered the shock fronts repetitively at 6–9 km s⁻¹, as required of a regenerative source.

Once Jupiter had formed, it would have continued to drive shock fronts in the inner nebula. And for as long as the gas and dust of the inner nebula remained, which was probably a few million years or less (J. Hester, Arizona State Univ.), chondrule-forming heating events would have kept occurring.

The picture of a slightly unstable nebular disk, in which large-scale transport of solids and gas can occur over timescales of thousands of years¹⁰, also explains the rapid mixing and transport of small lumps of solids that has led to the presence of both chondrules and CAIs in chondritic meteorites. In addition, solids would be forced by gas drag to migrate towards the centres of spiral arms. This explains why some of the CAIs and the first chondrules remained suspended in the

* Workshop on Chondrites and the Protoplanetary Disk, Kauai, Hawaii, USA, 8–11 November 2004 (www.lpi.usra.edu/meetings/chondrites2004).

nebula and were subject to repeated episodes of shock heating for a million years or so.

The Kauai workshop took place 20 years after another meeting in Hawaii on the origin of the Moon. That meeting is now recognized as the turning point in acceptance of the idea that the Moon formed following a giant impact of a Mars-sized body with Earth. The 2004 meeting may likewise take its place in history — as the time and place when it was explained how chondrules and CAIs were thermally processed by shock fronts in the solar nebula. ■

Alan P. Boss is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington,

5241 Broad Branch Road NW, Washington DC 20015-1305, USA.

e-mail: boss@dtm.ciw.edu

1. Sorby, H. C. *Nature* **15**, 495–498 (1877).
2. Hewins, R. H. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 3–9 (Cambridge Univ. Press, 1996).
3. Desch, S. J. & Connolly, H. C. Jr *Meteor. Planet. Sci.* **37**, 83–207 (2002).
4. Ciesla, F. J. & Hood, L. L. *Icarus* **158**, 281–293 (2002).
5. Halliday, A. *Nature* **431**, 253–254 (2004).
6. Bizarro, M., Baker, J. A. & Haack, H. *Nature* **431**, 275–277 (2004).
7. Amelin, Y., Krot, A. N., Hutcheon, I. D. & Ulyanov, A. A. *Science* **297**, 1678–1683 (2002).
8. Inaba, S., Wetherill, G. W. & Ikoma, M. *Icarus* **166**, 46–62 (2003).
9. Wood, J. A. *Meteor. Planet. Sci.* **31**, 641–645 (1996).
10. Boss, A. P. *Astrophys. J.* **616**, 1265–1277 (2004).

Diabetes

Outfoxing insulin resistance?

Marc Montminy and Seung-Hoi Koo

Resistance to insulin predisposes people to diabetes; it is characterized by increased storage of fats and a failure to stop glucose synthesis. The molecular underpinnings of these effects have been uncovered.

Type II diabetes affects some 5% of adults in most developed countries, and a far higher proportion of the population exhibits insulin resistance, a condition that predisposes individuals to diabetes. The mechanisms leading to insulin resistance are unclear, although the abnormal accumulation of certain fats in the liver (hepatic steatosis) is a contributing factor. On page 1027 of this issue, Stoffel and colleagues¹ show that the inactivation of a protein called Foxa2 promotes steatosis and contributes to the development of diabetes in insulin-resistant animals. The results have implications for the design of new drugs to treat insulin resistance and diabetes.

Energy balance in mammals resembles today's hybrid cars — we use a variable mix of glucose and fat as energy substrates, depending on food intake. During waking and feeding hours, we use glucose as an efficient source of energy, and under fasting conditions, during sleep for example, we burn primarily fat. Fat stores, called triglycerides, are converted to circulating fatty acids, and are further broken down in a process known as fatty-acid oxidation. In the fasting state, the liver also maintains normal circulating levels of glucose (which is essential for brain function) by synthesizing glucose anew, in a process known as gluconeogenesis.

The capacity of the liver to synthesize glucose and burn fat is governed by a set of ignition switches, called transcription factors, that operate in the nucleus to turn genes on and off². These switches respond to changes in circulating hormones — principally insulin and glucagon — that allow the liver cell to change gears between feeding and

fasting metabolism. In response to feeding, insulin triggers the activation of a cascade of proteins inside the cell, each transmitting the feeding signal to the next through a chemical modification known as phosphorylation. In insulin-resistant states, the orderly phosphorylation of certain proteins in response to insulin is damaged, preventing insulin from correctly regulating glucose and fat metabolism³. As a consequence, insulin-resistant individuals exhibit hyperglycaemia (high blood glucose), partly because of elevated gluconeogenesis in the liver.

Despite this inability to inhibit glucose production in diabetes, insulin still seems to be capable of shutting off the switch that normally promotes fat burning (fatty-acid oxidation) during fasting. This phenomenon, known as mixed insulin resistance, implies that the insulin signal is transmitted preferentially to the fatty-acid-oxidation switch rather than the glucose switch, the unfortunate consequence being that insulin-resistant individuals are not only hyperglycaemic but also accumulate triglycerides in the liver rather than breaking them down.

In their work, Stoffel and colleagues¹ test the notion that distinct switches control glucose and fat metabolism in the liver, and that one of these switches might be far more sensitive to insulin than the other. They show that the protein Foxa2, a member of the 'forkhead' family of transcription factors, is a key switch that regulates fatty-acid breakdown in the liver during fasting. Foxa2 resembles another forkhead-family member called Foxo1, which is well known to promote gluconeogenesis in liver in the fasted state³. Feeding inactivates both Foxo1 and Foxa2 switches through phosphorylation —

but the surprise, and the central point, of this paper¹ is that Foxa2 is far more sensitive to insulin signals than is Foxo1, and consequently Foxa2 is turned off even in insulin-resistant states, whereas Foxo1 is not.

Why is Foxa2 so much more sensitive to insulin? The answer is still murky, but the authors point to one possible explanation. Under feeding conditions, insulin stimulates two separate relay systems in liver cells that are distinguished by two key components: the insulin-receptor substrates IRS1 and IRS2 (ref. 4; Fig. 1). Stoffel and colleagues found that the Foxo1 switch is turned off only by the IRS2 relay system, but Foxa2 can be turned off by both IRS1 and IRS2. Thus, by virtue of its sensitivity to both IRS1 and IRS2 signals, Foxa2 may be turned off by insulin more readily than Foxo1.

Levels of circulating insulin are chronically elevated in insulin resistance, in part because of the abnormal gluconeogenesis in the liver, leading the authors to suggest that this hyperinsulinaemia causes Foxa2 to remain inactive, even during fasting periods. As a consequence of this block in fatty-acid oxidation, the liver begins to accumulate triglycerides. To test this model, Stoffel and colleagues used an altered Foxa2 with a mutation (denoted T156A) that blocks the phosphorylation of Foxa2 in response to insulin. Remarkably, when the mutant Foxa2 was

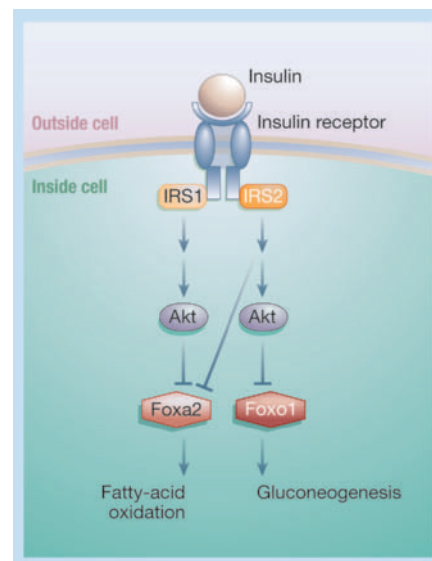


Figure 1 Model for insulin signalling in the liver. After a meal, insulin binds to its receptor in liver, activating two major pathways (one involving insulin-receptor substrate-1, IRS1, the other involving IRS2). Akt is a key enzyme in both signalling relays. The pathways shut down the synthesis of glucose (gluconeogenesis) and the burning of stored fats (fatty-acid oxidation) by phosphorylating Foxo1 and Foxa2. Stoffel and colleagues¹ now show that Foxo1 is phosphorylated only through the IRS2 pathway, whereas Foxa2 is phosphorylated in response to both IRS1 and IRS2, accounting for its enhanced sensitivity to insulin signals.

introduced into the livers of diabetic mice, it not only reversed hepatic steatosis but also improved insulin sensitivity. The results potentially explain how insulin can have different effects on glucose and fat metabolism, and provide clues to how hepatic steatosis develops and leads to diabetes.

Although this study brings up numerous questions, perhaps the most intriguing of these concerns the mechanism by which insulin shuts down Foxa2 preferentially over Foxo1. Presumably this difference reflects the involvement of certain components in the IRS1 and IRS2 pathways that associate with and favour Foxa2 phosphorylation. Identifying these components will be impor-

tant in understanding the basis of insulin selectivity. Regardless, the ability of the mutant Foxa2 to reverse hepatic steatosis provides a glimpse into potential new therapies for hepatic steatosis, and might one day provide clinicians with new measures to prevent the progression from insulin resistance to diabetes. ■

Marc Montminy and Seung-Hoi Koo are at The Salk Institute for Biological Studies, La Jolla, California 92037, USA.
e-mail: montminy@salk.edu

1. Wolfrum, C., Asilmaz, E., Luca, E., Friedman, J. M. & Stoffel, M. *Nature* **432**, 1027–1032 (2004).
2. Spiegelman, B. M. & Heinrich, R. *Cell* **119**, 157–167 (2004).
3. Saltiel, A. & Kahn, C. R. *Nature* **414**, 799–806 (2001).
4. White, M. *Mol. Cell. Biochem.* **182**, 3–11 (1998).

Applied physics

Nanotube antennas

M. S. Dresselhaus

An antenna array that is metres high and wide can detect and transmit radio waves. This effect has now been demonstrated at much smaller electromagnetic wavelengths in a nanoscale array of carbon nanotubes.

In *Applied Physics Letters*, Wang *et al.*¹ show in a clear way that an array of aligned carbon nanotubes can behave as an electromagnetic antenna. Their practical demonstration not only confirms a predicted effect but also points to its use in practical devices.

Antennas are familiar as detectors and transmitters of radio waves. But, depending on their dimensions, antennas can receive different wavelengths — radio, optical, microwave, and so on. All antennas have two major properties. First, their response varies with the polarization of the incoming radiation (polarization means that the electric field of the radiation has a particular orientation). Transmission is weakest if the plane of polarization is at 90° to the antenna's long axis: the polarization effect. And second, their response varies with their length, being strongest when the length is a multiple of half of the wavelength (λ) of the radiation (0.5 λ , 1.0 λ , 1.5 λ , and so on): the antenna-length effect.

The likely polarization and antenna properties of carbon nanotubes were recognized from a theoretical standpoint² soon after the first experimental synthesis^{3,4} of single-wall carbon nanotubes in 1993. Conceptually, single-wall carbon nanotubes (SWCNTs) can be considered to be formed by the rolling of a single layer of graphite (called a graphene layer) into a seamless cylinder. A multiwall carbon nanotube (MWCNT) can similarly be considered to be a coaxial assembly of cylinders of SWCNTs, like a Russian doll, one within another; the separation between tubes is about equal to that between the layers in natural graphite.

Hence, nanotubes are one-dimensional objects with a well-defined direction along the nanotube axis that is analogous to the in-plane directions of graphite.

Wang *et al.*¹ performed their experiments on random arrays of MWCNTs, aligned along their long axes and looking, at the nanoscale, like a dense forest of trees growing up from a silicon substrate. Each MWCNT in their nanotube array behaves effectively as a metallic rod about 50 nm in diameter and 200–1,000 nm in length (although within each array the MWCNTs are about the same length).

The particular novelty of this work is the clear and direct demonstration of the antenna-length effect. Using visible light, Wang *et al.* show that maxima occur in the amount of reflected light when the average length of the nanotubes in an array is a half-integral multiple of the wavelength of the incident light. They also provide a vivid demonstration of the polarization effect by comparing how much polarized light is reflected from a nanotube array and how much from a highly reflective metal surface positioned alongside the nanotube forest. For the metal surface, the electric-field vector must be in the plane of the metal surface for the light to be maximally reflected. In contrast, reflection from the nanotube arrays is strongest when the electric-field vector of the polarized light falls along the axis of the nanotubes (normal to the substrate).

These properties of nanotube arrays are related to the special electromagnetic behaviour of layered graphite, which has a strongly anisotropic 'skin depth'. The skin depth is the characteristic distance of penetration of

an electromagnetic wave into a material. If the incident radiation is polarized parallel to the plane of the graphite layers, the skin depth is short and the light is strongly absorbed. But if the polarization is normal to the plane of the layers, the penetration depth of the light is more than ten times larger and the absorption is very much weaker. This anisotropy in the electromagnetic properties of graphite was commonly exploited in the construction of infrared and microwave polarizers, before the development of Polaroid film in the 1960s.

Such polarization effects are useful in fundamental studies of the optical absorption, reflection and emission of carbon nanotubes. The intensity of the scattered light is proportional to the light's absorption and emission^{5–7} and is thus a sensitive probe of optical processes in carbon nanotubes. Furthermore, polarization has a major role in determining the right-handedness or left-handedness of a carbon nanotube⁸.

As Wang *et al.*¹ note, polarization effects in carbon nanotubes have been observed by several groups, in experiments on bundles of MWCNTs⁹ and of SWCNTs¹⁰, individual metallic SWCNTs¹¹ and individual semiconducting¹² SWCNTs, and on very-small-diameter (0.4 nm) SWCNTs in a zeolite template¹³. By performing polarization studies on an individual SWCNT, total suppression of the electric field normal to the nanotube axis could be demonstrated, and the dipole pattern of an individual nanotube could be studied — both when the tube was well isolated from other SWCNTs and when it interacted with a nearby SWCNT¹².

The work by Wang *et al.*¹ suggests a host of applications in optoelectronics, for example infrared polarizers or polarization detectors, using arrays in which the MWCNTs have a range of lengths, and wavelength-specific detectors, for which the MWCNTs are fabricated to all have the same desired length. MWCNTs are preferable to SWCNTs for their electrical metallicity and greater mechanical robustness, and their random spacing in the arrays suppresses interference effects.

The preparation of the MWCNT arrays could easily be extended to a high-throughput process with the use of established technology. Over the decade-long history of carbon nanotubes there has been substantial investment by the private sector in nanotube research and development: the technology devised by Wang *et al.*¹ may well spawn optoelectronic applications of considerable commercial significance. ■

M. S. Dresselhaus is in the Department of Physics and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: millie@mgn.mit.edu

1. Wang, Y. *et al.* *Appl. Phys. Lett.* **85**, 2607–2609 (2004).
2. Ajiki, H. & Ando, T. *Physica B Cond. Matt.* **201**, 349–352 (1994).
3. Iijima, S. & Ichihashi, T. *Nature* **363**, 603–605 (1993).
4. Bethune, D. S. *et al.* *Nature* **363**, 605–607 (1993).



100 YEARS AGO

Mankind in the Making, Anticipations and The Food of the Gods. By H. G. Wells. Three books of his more especially claim to forecast the future of our race... [M]en of energy — men of science, engineers, doctors, and so forth — will shape policy and administration. The result will be marvellous efficiency, such as is rarely if ever seen now. There will be no king. Monarchy will have given place to the New Republic. Royalty is connected with all things out of date, with aristocratic privileges, ridiculous costumes and decoration. Therefore it must go... The class that supplies unskilled labour, the old servile class, will tend to disappear. The invention of machines capable of performing more cheaply all the work that has hitherto fallen to the unskilled will make such men unnecessary. Peasant proprietors and all small land-holders must pass away. They represent stagnation, and there is room only for go-ahead adaptable people. Those who fail to adapt will fall into the abyss, the great sink in which wallow all those who are unfitted for the new conditions. The people of the abyss are to be encouraged to extinguish themselves, to practise what would commonly be called vice without offspring resulting.

From *Nature* 29 December 1904.

50 YEARS AGO

"Commonwealth Oceanographic Conference." It arose from arguments that the change of emphasis in oceanography, from exploration and survey to research directed towards the precise understanding of the basic physical and biological processes, makes it necessary to devote more effort to theoretical and experimental work... The only practical thing is for oceanographers to learn how the water movements are governed by atmospheric pressure, wind and climate so that they can use the meteorological data, supplemented every now and then by their own upper- and lower-water observations... It would no doubt be wrong to suggest that oceanographers took up the subject because they found the mathematical and physical sciences too difficult and unattractive; but they have been rather slow in applying the precise techniques of these sciences to marine research... To increase our knowledge the subject must be made attractive to men who do not mind facing up to the difficulties of fluid mechanics.

From *Nature* 25 December 1954.

5. Saito, R., Dresselhaus, G. & Dresselhaus, M. S. *Physical Properties of Carbon Nanotubes* (Imperial College Press, London, 1998).
6. Dresselhaus, M. S. & Eklund, P. C. *Adv. Physics* **49**, 705–814 (2000).
7. Dresselhaus, M. S., Dresselhaus, G., Jorio, A., Souza Filho, A. G. & Saito, R. *Carbon* **40**, 2043–2061 (2002).
8. Samsonidze, G. G. *et al. Phys. Rev. B* **69**, 205402 (2004).
9. Rao, A. M. *et al. Phys. Rev. Lett.* **84**, 1820–1823 (2000).
10. Jorio, A. *et al. Phys. Rev. Lett.* **85**, 2617–2620 (2000).
11. Duesberg, G. S., Loa, I., Burghard, M., Syassen, K. & Roth, S. *Phys. Rev. Lett.* **85**, 5436–5439 (2000).
12. Jorio, A. *et al. Phys. Rev. B* **65**, R121402 (2002).
13. Sun, H. D., Tang, Z. K., Chen, J. & Li, G. *Solid State Commun.* **109**, 365–369 (1999).

Molecular biology

Hairpins at split ends in DNA

Marjorie A. Oettinger

What do changing colours in corn kernels, mutations in houseflies and the variability of antibodies and of a T cell's antigen receptors in the vertebrate immune system have in common? A great deal, it turns out.

There once would have seemed to be little similarity between the random movement of transposable genetic elements within a fly, or a corn plant, or a flower petal, and the carefully orchestrated events by which antigen-receptor genes are assembled. The jumping of 'transposons' from one site to another in a genome can alter gene expression, giving rise to mutations that are seen as, for example, a colour change. By contrast, in the second reaction, a defined series of breakage and rejoining events within a chromosome brings together sequence elements in new combinations, for the express purpose of generating diversity in the receptors (antibodies and T-cell receptors) that detect foreign antigens. Nonetheless, a case for a mechanistic connection has been building for some time, and on page 995 of this issue, Zhou and colleagues¹ reveal a remarkable similarity in the steps by which the DNA is broken — a similarity that may extend to the structures of the enzymes responsible for this cleavage.

Transposition by the *hAT* family of mobile DNA elements (named after the *hobo* transposon from fruitflies, the maize *Ac/Ds* elements made famous by Barbara McClintock, and the *Tam* elements of snapdragon) involves the excision and complete removal of the DNA sequence of the transposable element from one site in the genome. The excised DNA sequence then jumps into another site. These reactions are mediated by a transposase — an enzyme encoded by the transposable element itself. The removal of the element creates breaks in the DNA that must be repaired to maintain the integrity of the genome. Tell-tale inverted duplications of a few base pairs of DNA are often found at the excision sites².

Meanwhile, in the V(D)J recombination reaction that assembles antigen-receptor genes, the RAG1 and RAG2 enzymes break both strands of DNA precisely at the border between protein-coding gene segments and their flanking recombination signal sequences. When the coding segments are rejoined in

new combinations, inverted repeats are also sometimes found at the junction².

The presence of inverted repeats at both excision sites and coding joints led to speculation that the transposition of *hAT* elements and V(D)J recombination might involve broken DNA intermediates of the same unusual structure: covalently sealed hairpins, in which the two separate strands of the DNA helix are joined together at the tip^{2,3}. If such structures were generated during breakage, and then reopened off-centre before repair, inverted repeats would be generated.

When hairpin intermediates were first proposed for plant transposons, it was suggested that two separate single-strand breaks were made in the DNA, one on the top strand and one on the bottom, followed by a ligation step that joined the broken ends together². But when hairpins were first physically detected during V(D)J recombination, it became clear that they are made by a more direct route⁴. Biochemical analysis with purified RAG1 and RAG2 proteins revealed that the hairpins arise as a necessary part of V(D)J cleavage. In a first step, a single-strand nick is introduced into the DNA, precisely at the border between coding and signal sequences (Fig. 1a). This single nick leaves a chemically reactive hydroxyl group at the 3' end of the top strand, which then attacks the opposite DNA strand in a transesterification reaction, yielding a hairpinned coding flank and a blunt, 5'-phosphorylated signal end.

As V(D)J recombination and transposition have been studied and compared, the parallels between them have mounted⁵, raising the question of whether the vertebrate immune system developed its unusual mechanism of generating diversity from a primordial transposon. In this case, it would be the excision step that was harnessed, not the step in which the released element jumps into another site. The DNA removed from the chromosome during V(D)J recombination is not moved to another site; rather, the ends are usually joined together to form a circular piece that is eventually lost from the

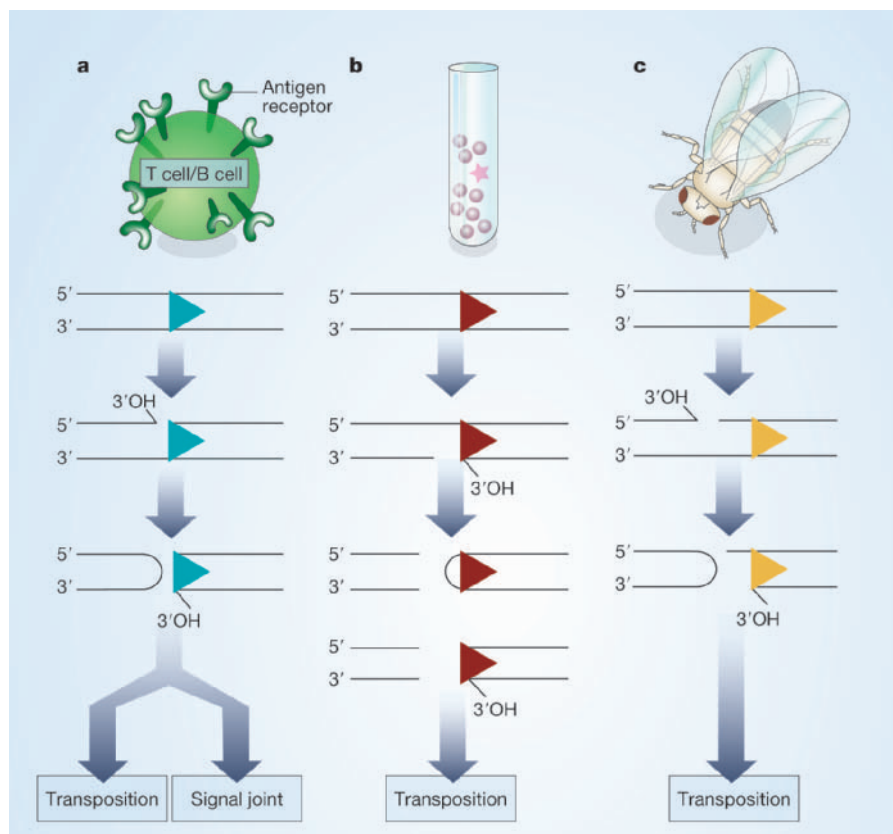


Figure 1 Hairpin formation in transposition and V(D)J recombination. The DNA-breakage steps in a, V(D)J recombination, which generates diversity in antigen-receptor genes; b, the random transposition of the bacterial Tn10 or Tn5 genetic elements, which can lead to altered gene expression (represented by the change from circles to stars); or c, the transposition of *Hermes* elements from houseflies, as just deduced by Zhou *et al.*¹. a, c, In the first steps of *Hermes*- and RAG-mediated cleavage, a nick is introduced on the top strand of the DNA, either one nucleotide 5' to the end of the transposition element (yellow triangle) or precisely at the border between the protein-coding sequence and the recombination signal sequence (blue triangle). b, By contrast, Tn5 or Tn10 transposition is initiated by a nick on the bottom strand. A hairpin intermediate is generated on the flanking DNA in the *Hermes*- and RAG-mediated reactions, and on the end of the transposon sequence in Tn5/Tn10 transposition. Following DNA breakage, the excised elements can be transposed to a new site (b,c). In V(D)J recombination (a), the signal ends are normally joined to each other in a 'signal joint', making them unavailable for transposition.

cell. This is not unknown in the transposon world, but with other transposons it is a rare event, whereas with V(D)J recombination it is the main pathway.

Despite these distinctions, there are similarities, and striking ones⁷. The unusual genomic arrangement of the RAG genes, found side-by-side in the genome and lacking introns (non-coding sections) within their protein-coding regions, is more like that found for bacterial or viral genes than for genes in higher eukaryotes. The RAGs have the ability to act as a transposase *in vitro*^{6,7} and in yeast⁸, even if they do so rarely in developing immune cells⁹. And some other transposable elements do use hairpin intermediates^{10,11}. However, the arrangement is different: the hairpin is formed on the ends of the transposable element rather than on the flanking DNA (Fig. 1b). Is there a transposon that does it like the RAGs?

Zhou and colleagues¹ have now provided the missing link, showing that the housefly

Hermes element, a member of the *hAT* family, does leave hairpinned DNA behind when it cuts itself out of DNA *in vitro*, as suggested². As with the breaks generated by RAGs, the hairpin is on the DNA flanking the excised element (Fig. 1c), and arises from a top-strand nick that is converted directly to a hairpin (presumably by the same type of one-step transesterification, but this awaits demonstration).

And the similarities don't end there. Like the RAGs and members of the retroviral integrase superfamily¹², the *Hermes* protein contains a catalytic triad of acidic amino acids (a DDE motif) that is crucial for the protein's activity. Hitherto, RAG1 was the outlier in the superfamily, with a far greater separation in its amino-acid sequence between the second D (aspartate) and E (glutamate) residues than has been found in other family members. But *Hermes* shares this wider separation, hinting that there is some connection between this spacing and

the generation of a hairpin on the flanking DNA. Zhou *et al.* also suggest that the DDE motif of *hAT* transposases and RAG1 lies in an RNaseH-like fold, which is common to members of the retroviral integrase superfamily¹³. That points to still closer connections between all these enzymes.

But why a hairpin? The prevalence of this intermediate suggests some biological importance, but precisely what that is remains unclear. One possibility is that a hairpin intermediate provides a molecular solution to the problem of breaking DNA by using an enzyme with a single active site. It has been suggested that the structural alterations required to achieve the chemistry of the first hydrolysis step and a subsequent intramolecular transesterification pose less difficulty than catalysing two separate hydrolysis events on opposing strands¹⁴. Another possibility is that the hairpin somehow aids in the resolution of the break, perhaps attracting appropriate repair factors or ensuring the proper rejoining of paired ends. As the details of how the hairpin is generated and processed are uncovered, answers to this question may be obtained.

Other questions include: what determines whether the excised fragment is transposed or the ends are harmlessly circularized? Will *hAT* elements be excised in mammalian cells and the ends rejoined, or have RAG-specific mechanisms evolved to promote V(D)J recombination and suppress RAG-mediated transposition? Are the same host factors required to process the hairpin DNA and heal the broken chromosome in both cases? Are there structural features of the *Hermes* and RAG proteins that explain why the hairpin is made on the flanking DNA rather than the recognition sequence? Through the efforts of Zhou and colleagues, we have a new opportunity to advance our understanding of two distinct—but remarkably similar—processes.

Marjorie A. Oettinger is in the Department of Genetics, Harvard Medical School, and the Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. e-mail: oettinger@frodo.mgh.harvard.edu

1. Zhou, L. *et al.* *Nature* **432**, 995–1001 (2004).
2. Coen, E. S., Robbins, T. P., Almeida, J., Hudson, A. & Carpenter, R. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 413–436 (Am. Soc. Microbiol., Washington DC, 1989).
3. Lewis, S. M. *Adv. Immunol.* **56**, 27–150 (1994).
4. Gellert, M. *Adv. Immunol.* **64**, 39–64 (1997).
5. Brandt, V. L. & Roth, D. B. *Immunol. Rev.* **200**, 249–260 (2004).
6. Hiom, K., Melek, M. & Gellert, M. *Cell* **94**, 463–470 (1998).
7. Agrawal, A., Eastman, G. M. & Schatz, D. G. *Nature* **394**, 744–751 (1998).
8. Clatworthy, A. E., Valencia, M. A., Haber, J. E. & Oettinger, M. A. *Mol. Cell* **12**, 489–499 (2003).
9. Messier, T. L. *et al.* *EMBO J.* **22**, 1381–1388 (2003).
10. Kennedy, A. K., Guhathakurta, A., Kleckner, N. & Haniford, D. B. *Cell* **95**, 125–134 (1998).
11. Bhasin, A., Goryshin, I. Y. & Reznikoff, W. S. *J. Biol. Chem.* **274**, 37021–37029 (1999).
12. Haren, L., Ton-Hoang, B. & Chandler, M. *Annu. Rev. Microbiol.* **53**, 245–281 (1999).
13. Rice, P. A. & Baker, T. A. *Nature Struct. Biol.* **8**, 302–307 (2001).
14. Kennedy, A. K., Haniford, D. B. & Mizuuchi, K. *Cell* **101**, 295–305 (2000).

Climate change

The cloud conundrum

Joyce E. Penner

One of the great uncertainties in projecting global warming is accounting for the effects of small particles in Earth's atmosphere. Progress is nonetheless being made with this fiendishly complex problem.

Generally accepted climate projections for the year 2100, compared with today, predict a global average temperature increase ranging from around 5.8 °C to a more benign, but still worrisome, 1.4 °C. Which of these futures awaits us depends, in part, on aerosols — tiny particles in the atmosphere — and their effect on clouds. Ackerman *et al.* (page 1014 of this issue)¹ conclude that the role of these particles in increasing the water content of clouds, and so cloud reflectivity, is smaller than previously thought.

As a consequence of air pollution, the mass of small particles in the atmosphere is about 40% larger than it would be in a pristine atmosphere². These particles mainly reflect solar radiation, a direct effect on climate that has received considerable attention over the past 15 years. But when clouds form, these tiny particles also act as seeds for water condensation. When a cloud forms in a region containing a large number of particles, it will have higher concentrations of smaller droplets if the liquid water in the cloud is constant. The smaller droplets have a higher surface area; so the cloud is brighter and able to reflect more solar radiation, thereby cooling the climate. This is the so-called first indirect effect of aerosols on clouds.

But climate scientists have also explored the consequences of smaller droplets on precipitation. Smaller droplets are less likely to collide with each other and form precipitation. This change in 'precipitation efficiency'

has been termed the second indirect effect, and has been calculated to increase both the total cloud cover and the total amount of liquid that is held within clouds. These two effects increase the total reflected solar radiation even more and lead to a conundrum. Because some climate models are highly sensitive to this change, they predict a total cooling effect that is larger than the warming by greenhouse gases: but the net effect of greenhouse-gas and aerosol changes over the past 100 years is observed to be a warming.

The conundrum may be solved in one of three ways. Some particles, especially the 'black carbon' produced by biomass and fossil-fuel burning, can absorb radiation rather than reflect it, and so counteract cooling. Perhaps this effect has been underestimated: if these particles are much more abundant than thought³, or if they cause more absorption of solar radiation after deposition on snow⁴ (which is normally highly reflective), they may be counteracting the effects of particles on precipitation efficiency to a greater extent than expected.

Or perhaps the effects of particles on 'ice clouds' — clouds high in the atmosphere — have been underestimated. Because high clouds tend to absorb more energy in the form of thermal radiation than they reflect in the form of solar radiation, an increase in ice-cloud amount and particle number would warm the climate. Almost no work has addressed the effect of air pollution on ice clouds; but if increased pollution also causes

increases in ice particles within high clouds, it might balance the effect of increased air pollution on low, warm clouds.

A third possibility is that those climate models in which clouds are very sensitive to changes in particle concentrations have got it wrong. Ackerman *et al.*¹ show that this might indeed be the case, which is perhaps not unexpected. Because climate models are typically run at a resolution of only 250 km, they cannot resolve individual clouds, and must parametrize the effects of particles on clouds — that is, they invoke comparatively crude equations of cloud physics, rather than simulate real clouds.

Ackerman *et al.*¹ used a large eddy-simulating, cloud-resolving model to examine the effects of increased particle concentration on low-lying stratocumulus clouds. They picked a number of different cases that have been studied in the past and show that an increase in the number of particles does not lead to an increase of total liquid — unless, that is, the overlying air in the region is humid. The reason is that when precipitation decreases as a result of increased particle concentrations, the mixing of air from above the cloud with that beneath it increases. If the air above the cloudy layer is dry, the increased mixing leads to less liquid water in the cloud. But if the air above the cloud is humid, then the mixing resupplies the cloud with moisture. Ackerman and colleagues' new results restrict the number of regions that may be affected by the second indirect effect to a smaller percentage of the globe, and so mean that the total impact of that effect may be much smaller than previously thought.

If the first indirect effect — cloud brightness — is as large as some models predict, aerosols are still acting strongly to increase reflection of solar radiation, and so are masking the current effect of greenhouse gases to a large extent. If so, then the net radiative forcing that is the cause of the 0.6 °C average

temperature increase of the past 100 years must be small, and climate models must be much more sensitive to this small difference if they are to agree with past observations. The future, then, might be more at the upper range of climate projections. But if aerosols do not increase low-level cloud brightness as much as some models have it, then greenhouse gases may be only slightly masked by aerosol-induced cloud changes, and projections of future climate might follow the more benign path.

How fast the world needs to address the issue of greenhouse-gas warming still depends on which of these two paths is correct. We still have the conundrum. But it

is now less daunting than it was, given that the results of Ackerman *et al.* mean we can rule out some of the large effects of aerosols on cloud water content. ■

Joyce E. Penner is in the Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, 2455 Hayward, Ann Arbor, Michigan 48109-2143, USA.
e-mail: penner@umich.edu

1. Ackerman, A. S., Kirkpatrick, M. P., Stevens, D. E. & Toon, O. B. *Nature* **432**, 1014–1017 (2004).
2. Penner, J. E. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 289–348 (Cambridge Univ. Press, 2001).
3. Sato, M. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 6319–6324 (2003).
4. Hansen, J. & Nazarenko, L. *Proc. Natl Acad. Sci. USA* **101**, 423–428 (2003).

Developmental biology

Survival by self-digestion

Nathaniel Heintz

Mammals face a problem just after birth: they are no longer nourished through the placenta, but suckling has not yet begun. How do they survive? Digestion of the animal's own cells could be the answer.

All organisms must adapt to environmental stresses. One of the most beautiful examples of a response to stress is autophagy — a process first described in electron microscopic studies in the kidneys of newborn mice¹, and now known to be one of the major pathways for the degradation of long-lived proteins and cellular organelles^{2,3}. It has become clear that autophagy is activated in a variety of circumstances in multicellular organisms, and thus must be regulated by diverse stimuli in different cell types. But our knowledge of the biological roles of autophagy in mammals is quite primitive. On page 1032 of this issue, however, Kuma *et al.*⁴ provide a fascinating insight into the mammalian response to birth and the biological benefits of autophagy.

In cells, nutrient deprivation and other stresses elicit a complex series of orchestrated steps that result in the cells enveloping portions of their cytoplasm, delivering them to enzyme-packed organelles known as lysosomes for degradation and recycling (hence the name autophagy, or self-eating)^{2,3}. Over the past decade, genetic analysis in yeasts and morphological studies in a wide variety of species have defined the molecular mechanisms of autophagy and shown its induction under many developmental and pathological situations. The logic for the pathway that has been revealed by the genetic and molecular dissection is very pleasing — in response to nutrient deprivation, the autophagy pathway scavenges intracellular constituents to supply vital components until conditions improve.

To investigate the importance of this pathway in mammals, the Mizushima laboratory

has previously⁵ made use of a fusion of green fluorescent protein (GFP) and LC3, the mammalian equivalent of Atg8 — an essential protein in the yeast autophagy pathway. With this fusion protein they could reveal the formation of autophagosomes, vesicular structures that are hallmarks of autophagy. Having introduced GFP–LC3 into young mice, the authors were able to map the induction of autophagy in response to food withdrawal in many tissues, by studying the formation of small fluorescent vesicles carrying the fluorescent fusion protein.

Now, the same group (Kuma *et al.*⁴) have turned their attention to the developmental roles of autophagy. To do so, they first used GFP–LC3 to study the formation of autophagosomes in newborn mice. They observed massive induction of autophagy in the heart muscle, the diaphragm, the lungs and the skin — all tissues that undergo sudden increases in energy expenditure or environmental exposure immediately following birth. The induction of autophagy was immediate and transient, reaching maximal levels only 3 to 6 hours after birth and declining to basal levels within a day or two. To explain this phenomenon, Kuma *et al.* postulated that the induction of autophagy following birth is required to provide energy before nursing begins.

To test this attractive idea, the authors generated mice that lack Atg5 — another protein required during an early step in the autophagy programme⁶. The Atg5-knockout animals were born in normal mendelian ratios but died during the first day after birth. Measurements of autophagy with the GFP–LC3 assay demonstrated that, as expected, the

formation of autophagosomes was blocked in the mutant mice. The Atg5-deficient mice also died earlier than wild-type mice when not being suckled; their survival could be extended by hand feeding; and plasma and tissue amino-acid concentrations in non-suckling mutant mice were normal at birth but very much reduced relative to controls 10 hours after delivery. Furthermore, measurements of energy status in the knockout pups indicated that energy production was severely depressed but could be restored by re-feeding.

These data show that autophagy is required to produce amino acids and to maintain energy levels in neonatal mice. Taken together with the severely low glucose and lipid levels that are generally evident in newborn mice, the death of the Atg5-knockout mice after birth provides strong support for the hypothesis that autophagy is essential for survival during the unique period experienced by mammals as they make the transition from transplacental nutrition to milk.

The idea that autophagy is important to provide nutrients for neonatal survival is very attractive. It provides a genetic link between the incisive studies of mechanisms of autophagy in individual cells, and studies documenting the induction of autophagy in tissues of fasting animals^{2,3}. It demonstrates that our understanding of the biological roles of autophagy in mammals will ultimately require precise genetic dissection of this pathway *in vivo*. And it raises several issues for further investigation.

For instance, under what other conditions can autophagy contribute to viability in mammals and other vertebrates? What is the mechanism through which autophagy, which is understood as an intracellular catabolic pathway (one that breaks complex molecules down), contributes to the maintenance of plasma amino-acid levels? Which tissues contribute to this process? How important is this mechanism in other organisms in maintaining energy levels during development or fasting? What is the anatomical basis for the suckling defect seen in Atg5-knockout mice, and how does that contribute to neonatal death? In the studies presented here, Kuma *et al.*⁴ make a powerful case that our understanding of the unique features of mammalian development can be enriched by genetic dissection of this beautiful and fundamental biological pathway. ■

Nathaniel Heintz is in the Laboratory of Molecular Biology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.
e-mail: heintz@mail.rockefeller.edu

1. Clark, S. L. *J. Biophys. Biochem. Cytol.* **3**, 349–362 (1957).
2. Levine, B. & Klionsky, D. J. *Dev. Cell* **6**, 463–477 (2004).
3. Kamada, Y., Sekito, T. & Ohsumi, Y. *Curr. Top. Microbiol. Immunol.* **279**, 73–84 (2004).
4. Kuma, A. *et al.* *Nature* **432**, 1032–1036 (2004).
5. Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *Mol. Cell. Biol.* **15**, 1101–1111 (2003).
6. Mizushima, N. *et al.* *J. Cell Biol.* **152**, 657–668 (2001).

No red cell is an island

James Palis

Red blood cells develop in the bone marrow in 'islands' nurtured by a central white blood cell. Work in mice shows that the retinoblastoma protein is crucial for these white cells to mature and form islands.

As adults, we make more than 2 million red blood cells (erythrocytes) in our bone marrow every second to replace those lost through normal attrition. The mammalian erythrocyte is unique among the blood cells of the animal kingdom because it discards its nucleus (enucleates) and all its internal organelles before entering the circulation. This maturation of red cells is aided by white blood cells called macrophages, and somehow requires the retinoblastoma protein (Rb). On page 1040 of this issue, Iavarone *et al.*¹ tease out some of the mechanisms involved. Surprisingly, it seems that Rb regulates the maturation of macrophages, which bind to immature nucleated erythrocytes (erythroblasts) to form 'islands' that nurture red-cell synthesis.

In the early 1940s, Giovanni Di Guglielmo observed that erythroblasts are associated with macrophages in the bone marrow. Subsequently, pioneering studies by Marcel Bessis² indicated that all red cells mature within 'erythroblast islands' consisting of a central macrophage that extends cytoplasmic protrusions to a ring of surrounding erythroblasts (Fig. 1). Such islands exist in adult bone marrow and also in fetal liver, where erythrocytes are produced before the bone marrow is fully developed. Bessis postulated that the central macrophage nurses erythroblasts by supplying nutrients and growth factors. The macrophages also gobble up the nuclei left behind by mature erythrocytes as they leave the island to enter the turbulent world of the circulation.

The Rb protein is a nuclear protein found in all cells and is key to the cell's decision to traverse the cell cycle and divide. Mice lacking Rb die at mid-gestation (embryonic day 14.5), having defects in the nervous system, placenta and eye lens, as well as deficiencies in the production of red blood cells. Specifically, erythroblasts in the livers of Rb-deficient fetuses mature abnormally and fail to enucleate^{3,4}.

The mechanism responsible for this Rb-dependent defect in red-cell maturation has remained controversial for more than a decade. Is the role of Rb intrinsic to maturing red cells, or is it extrinsic — that is, does Rb act through other cell types that provide a supportive microenvironment for erythroblasts? Red cells derived from Rb-deficient erythroid progenitor cells fail to differentiate normally when grown in culture⁵, suggesting that the erythroid defect is intrinsic to red cells. However, chimaeric mice composed of both normal and Rb-deficient cells generate normal erythrocytes that are derived from Rb-deficient cells⁶, suggesting instead that the defect is extrinsic to the red-cell lineage. The controversy over the site of action of Rb has been further fuelled by a study of Rb-deficient haematopoietic stem cells (that is, cells that can generate all the blood cells found in the body). When these mutant stem cells were transplanted into normal adult mice, the animals developed anaemia and their red cells failed to enucleate⁷, suggesting that Rb function is intrinsic to the red-cell lineage. However, these results could also be explained if the Rb protein has a key

role in another blood-cell type that supports erythroid maturation.

Iavarone *et al.*¹ now provide evidence that Rb-deficient fetuses have a severe defect in the macrophage lineage and fail to form functional erythroblast islands. Not only is there a marked reduction in mature macrophages in the liver of Rb-deficient fetuses, but the immature macrophages that are present cannot bind to erythroblasts. Using an ingenious erythroblast-macrophage co-culture system to reconstitute erythroblast islands *in vitro*, the authors find that Rb-null fetal erythroblasts can bind to normal macrophage cells and mature just as well as do normal erythroblasts in this system, suggesting that the erythroid defect is extrinsic to the red-cell lineage.

The authors also shed light on the role of Rb in macrophage differentiation. They had shown previously that the nuclear protein Id2 is bound by Rb and mediates some of the effects of Rb⁸. By genetically crossing mice lacking Id2 with mice lacking Rb, they find that the macrophage defect evident in Rb-null mice can be reversed by loss of Id2. They go on to show that Id2 can bind to the transcription factor PU.1, which is thought to be a critical regulator of genes expressed in macrophages, and that this interaction inhibits the transcription of macrophage-specific genes. These results are consistent with a model in which Rb binds to Id2 in the nucleus of immature macrophage cells, thus freeing PU.1 to increase the expression of genes involved in the development of macrophages into mature erythroblast-nursing cells.

Iavarone and colleagues' results are unlikely to be the final word in the erythroid-cell intrinsic-versus-extrinsic controversy. For instance, a recent report⁹ suggests that Rb-deficient red cells have a maturational defect that is cell-intrinsic: although Rb-deficient red-cell progenitors expand normally *in vitro*, when they are then placed in conditions that induce terminal red-cell maturation, they fail to exit the cell cycle normally and later show defective enucleation. Maybe the correct question is not whether the Rb-dependent maturation defect is exclusively intrinsic or extrinsic to red cells, but rather how much of the defect is attributable to each mechanism.

We have much still to learn about the interplay of red cells and macrophages in erythroblast islands. Little is known about the factors that promote their intimate contact or exactly how macrophages support erythroid maturation. However, these recent studies provide new insights into the functions of the Rb gene and the mechanisms governing normal erythrocyte and macrophage maturation.

James Palis is in the Department of Pediatrics and Center for Human Genetics and Molecular Pediatric Disease, University of Rochester Medical Center, Box 703, 601 Elmwood Avenue,

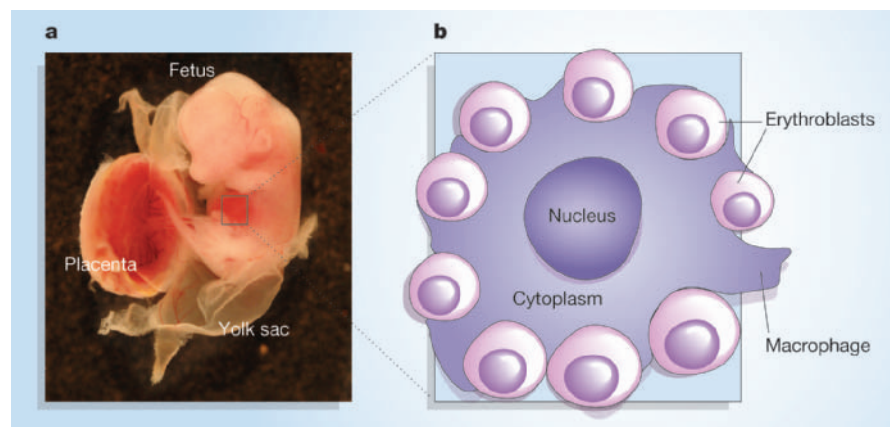


Figure 1 Erythroblast islands. a, A mouse fetus at embryonic day 13.5. Red blood cells are produced in the liver, which is bright red. b, Red blood cells mature in erythroblast islands, each of which has a central white blood cell, the macrophage, in contact with a ring of immature, nucleated red cells (the erythroblasts).

Rochester, New York 14642, USA.
e-mail: james_palis@urmc.rochester.edu

1. Iavarone, A. *et al. Nature* **432**, 1040–1045 (2004).
2. Bessis, M. *Rev. Hematol.* **13**, 8–12 (1958).
3. Lee, E. Y.-H. P. *et al. Nature* **359**, 288–294 (1992).

4. Clarke, A. R. *et al. Nature* **359**, 328–330 (1992).
5. Jacks, T. *et al. Nature* **359**, 295–300 (1992).
6. Williams, B. O. *et al. EMBO J.* **13**, 4251–4259 (1994).
7. Hu, N. *et al. Cancer Res.* **57**, 4123–4129 (1997).
8. Lasorella, A. *et al. Nature* **407**, 592–598 (2000).
9. Clark, A. J. *et al. Blood* **104**, 1324–1326 (2004).

Device physics

Microlever chilled out

Peter W. Milonni and Boris M. Chernobrod

New work shows how light might be used to cool a micrometre-size cantilevered mirror to the low temperatures required in physics experiments and applications.

Microfabricated cantilevers have permitted by far the most sensitive studies of small forces at tiny length scales. They are the basis of, among other things, the atomic force microscope and the magnetic-resonance force microscope. Microlevers are increasingly of interest as tools for the study of materials and the development of quantum microscopes for such fundamental investigations as detection of the presence of a single unpaired electron spin¹. To reduce thermal fluctuations — and therefore increase the sensitivity of these techniques — it is generally desirable to work at low temperatures; experiments in single-spin detection, for example, are typically performed at 1.6 K. Other possible applications, such as the detection of gravitational waves or the study of the quantum superposition of states of photons or macroscopic objects, also require very low temperatures. Whereas the laser cooling of atomic gases is now commonplace, progress in the laser cooling of macroscopic objects has been less dramatic. On page 1002 of this issue, H hberger Metzger and Karrai² report a promising step in this direction.

Their work builds on earlier studies of radiation forces acting on macroscopic objects, in particular a mechanical effect of radiation³ that is used to cool one of the mirrors of an optical resonator, using feedback control⁴. With a resonator consisting of two parallel and highly reflecting mirrors, the phase of the laser light that is reflected from the resonator varies markedly with changes in the separation of the mirrors. Measurement of the phase variation in this laser light allows detection of the thermal (brownian) motion of a mirror if it is sufficiently lightweight. It can also provide the signal for a feedback loop that controls the motion of the mirror, by adjusting the force, resulting from radiation pressure, that is exerted on it by a second laser beam: if the force is adjusted so that it opposes the motion of the mirror, this results in a reduction (cooling) of the mirror's brownian motion.

But the microlever cooling technique reported by H hberger Metzger and Karrai²

does not use a feedback loop or radiation pressure from a second laser. Instead, it relies on the force on a resonator mirror that is generated by the light inside the resonator. This force could be simply radiation pressure, or a photo-thermal stress resulting from the absorption of radiation by different parts of the mirror possessing different coefficients of thermal expansion (such as the bulk of the microlever and a metal coating). In either case, the force is proportional to the intensity of light stored in the cavity between the mirrors, and is greatest when the separation of the mirrors is such that the laser wavelength matches a cavity (Fabry–P rot) resonance.

Because the mirror separation is changed by the force, it is strongly coupled to the light intensity inside the cavity. Changes in this intensity do not occur instantaneously with mirror separation. Therefore, the delay in the response of the intensity to a change in the mirror separation leads to a force that can enhance or oppose the motion of a mirror, depending on whether the optical frequency is higher or lower than that of the cavity resonance. The motion of the mirror can be modelled using Newton's second law, having a total force that includes this intensity-dependent force, as well as a fluctuating thermal force. Such a model accounts accurately for H hberger Metzger and Karrai's results².

In their experiment, the resonator is formed, inside a vacuum, by two mirrors: one is a semi-transparent gold film across the end of an optical fibre, and the other is a gold-coated silicon microlever of dimensions $223\text{ }\mu\text{m} \times 22\text{ }\mu\text{m} \times 0.46\text{ }\mu\text{m}$ (Fig. 1). The separation of the mirrors is about $34\text{ }\mu\text{m}$. Light from a HeNe laser (with a wavelength of 633 nm) propagates from the fibre into the resonator; the light reflected from the resonator propagates back through the fibre and is analysed to deduce the thermal noise spectrum of the microlever displacement and therefore its temperature. In this way, the authors determined that the microlever temperature was reduced as the laser power increased, cooling from room temperature to a temperature of 18 K. This corresponds to a reduction of nearly a factor of 100 in

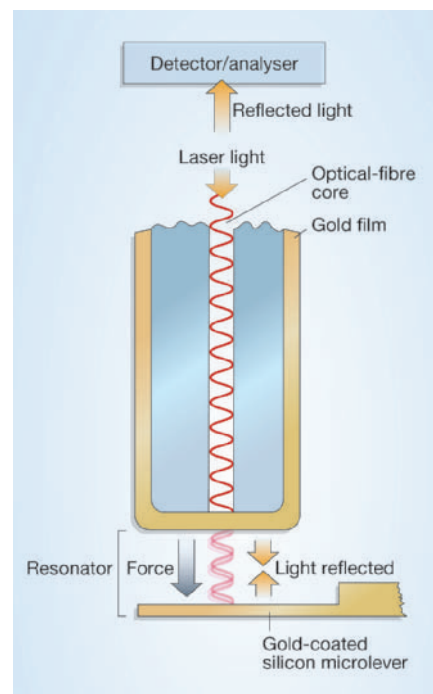


Figure 1 Cool for cavities. In the experimental set-up used by H hberger Metzger and Karrai², laser light passing through an optical fibre is reflected by the mirrors formed by a gold film on the tip of the fibre and a gold-coated silicon microlever. Inside this resonant cavity, the light exerts pressure on the microlever, through radiation pressure or through the forces induced by absorption of the light (a photo-thermal effect, which the authors² believe dominates here). The signal transmitted back through the optical fibre is used to deduce the thermal noise spectrum and therefore the reduction in temperature resulting from the light pressure inside the cavity.

the amplitude of the brownian motion.

The key ingredient for the cooling effect is a force that is delayed with respect to a change in the mirror separation. Because of the small mirror separation and low mirror reflectivities, the radiation pressure in this experiment changes almost instantaneously with fluctuations in the microlever displacement. This cooling effect is therefore attributed by the authors² to a photo-thermal (bolometric) force. The degree of cooling will be limited by the residual heating that results from optical absorption by the lever as well as the lever's fundamental quantum fluctuations; but, theoretically at least, cooling to sub-millikelvin temperatures is feasible with this technique.

Peter W. Milonni and Boris M. Chernobrod are in the Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. e-mail: pwm@lanl.gov

1. Rugar, D. *et al. Nature* **430**, 329–332 (2004).
2. H hberger Metzger, C. & Karrai, K. *Nature* **432**, 1002–1005 (2004).
3. Braginsky, V. B. & Manukin, A. B. *Sov. Phys. JETP* **25**, 653–655 (1967).
4. Cohadon, P. F., Heidmann, A. & Pinard, M. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).

Obituary

Yasutomi Nishizuka (1932–2004)

Ever since the discovery of hormones at the beginning of the last century, the cellular responses to these molecules have been high on the research agenda for biologists. During the 1970s and 1980s, answers began to emerge at the molecular level, most notably through the identification of key molecules and events that follow the binding of a hormone to receptors on the cell surface, and transfer the hormonal signal to the cell nucleus to elicit a response.

So it was that terms and processes that are now textbook material were unveiled. The activation of G proteins, which are associated with hormone receptors inside the cell, leads to the activation of protein kinase A. Activated G proteins can also trigger the metabolism of phospholipids in the cell membrane, particularly cleavage of phosphatidylinositol-4,5-bisphosphate (PIP₂) into inositol-1,4,5-trisphosphate (IP₃) and diacylglycerol, a molecule that in turn activates protein kinase C. The activated protein kinases A and C phosphorylate other proteins, initiating a cascade of phosphorylations that hand the activation signal from one protein to the next until it reaches the nucleus and stimulates the correct response.

Yasutomi Nishizuka, who died on 4 November, was one of the main figures in this biological revolution. In 1989, he was a co-winner of the Lasker award for the discovery of protein kinase C, along with Michael Berridge for the identification of IP₃ function, Alfred Gilman for the discovery of G proteins and their function, and Edwin Krebs for his pioneering work on protein phosphorylation. Gilman and Krebs later received the Nobel prize for their contributions.

Nishizuka became interested in protein phosphorylation when he visited the Rockefeller University in New York in 1964 and 1965 to work with Fritz Lipmann on how proteins are synthesized. During his stay, he purified proteins called elongation factors G and T, which aid the growth of the amino-acid chain, and elucidated the role of GTP in the process.

Nishizuka's time with Lipmann and his colleagues was comparatively brief, but it gave him a strong impetus to work on the biological function of heavily phosphorylated nuclear proteins. Lipmann had himself studied phosphorylated proteins in the 1930s, when he was a visiting scientist at the Rockefeller. But even in the 1960s, the regulatory roles of phosphate were not



Discoverer of key protein activation pathways

widely recognized, except for the conservation of bioenergy in the high-energy phosphate discovered by Lipmann, and the control of glycolysis by the phosphorylation of key proteins elucidated by Edmond Fischer and Krebs. So when Nishizuka returned to Kyoto University in 1965, he started work on an enzyme that phosphorylated other proteins, which later turned out to be protein kinase A.

In January 1969, Nishizuka was appointed professor of biochemistry at Kobe University School of Medicine. His new lab had belonged to the Department of Industrial Medicine, and he was surprised to find the space occupied by bicycles and a large bathtub, with no biochemical equipment or lab benches. He wasted no time in setting up the lab, however, and attracted several brilliant medical students to join him as postgraduates. The team began studying calcium-dependent protein phosphorylation. Nishizuka originally thought that a calcium-dependent protease triggered the kinase activity in brain cell extracts, but it soon became clear that the membrane phospholipid fraction was responsible, and the activator turned out to be diacylglycerol. Nishizuka coined the name protein kinase C for this kinase, and he

hypothesized that it could be linked with PIP₂ metabolism.

Nishizuka made another breakthrough in 1982, following an inspiring encounter with Monique Castagna, who visited Kobe from Paris. Castagna's interest was the tumour-promoting properties of phorbol esters, and the outcome of her collaboration with Nishizuka was the finding that these compounds activate protein kinase C continuously, thus causing uncontrolled cell growth.

Born into a medical family, Nishizuka had followed the lead of his older brother Yasuaki, a well-known pathologist, choosing to become a biochemist after studying medicine. Nishizuka began his scientific career in 1958 as the first graduate student of Osamu Hayaishi in the Department of Medical Chemistry at Kyoto University. Hayaishi had just returned to Japan from the US National Institutes of Health. At the time, Japanese biochemistry was dominated by chemistry-oriented research, mostly aimed at identifying new chemical compounds in metabolites, but Hayaishi introduced the practice of studying processes and reactions rather than just end products. Consequently, Nishizuka had a thorough grounding in enzymology, and the importance of considering the physiological function of molecules was imprinted on his science.

As a scientist, Nishizuka was very critical when considering data — I remember him as always being the last person to be convinced by new results. But he was otherwise warm-hearted and gentle, with a sense of humour. He revered the arts, particularly music (he played the piano) and painting. His artistic sense was evident in his lecture slides — quite an achievement, because for much of his career each letter had to be stencilled by hand.

On 17 January 1995, a huge earthquake struck Kobe, destroying several of the university buildings — and damaging Nishizuka's home. Only two weeks after this disaster, Nishizuka was named president of Kobe University, and in the ensuing six years of his term of office he rebuilt the university and re-established it as a competitive academic institution. At the time of his death he was serving as president of Hyogo Prefectural Centre for Adult Disease and on many government advisory committees. He will be remembered as a great and charismatic scientist.

Tasuku Honjo

*Tasuku Honjo is in the Department of Medical Chemistry and Molecular Biology, Kyoto University Faculty of Medicine, Kyoto 6068501, Japan.
e-mail: honjo@mfour.med.kyoto-u.ac.jp*

Biological chemistry

Ringing the changes with vancomycin

Angew. Chem. Int. Edn **43**, 6709–6713 (2004)

Vancomycin is a front-line antibiotic, but the only large-scale source is a natural product made by bacteria. Vancomycin analogues cannot be chemically synthesized in large quantities, but such an ability could be of great advantage in engineering new variants to combat the growing threat of drug-resistant bacteria. It might be possible to alter the substrate specificity of key proteins involved in vancomycin biosynthesis, and produce analogues that attack drug-resistant bacterial strains.

But this is no simple task: vancomycin biosynthesis involves many enzymatic steps, including three ring-forming reactions whose mechanism and substrate specificity are poorly understood. Katja Zerbe *et al.* have now performed one of these reactions *in vitro*, using an enzyme called OxyB to form the first ring in a simplified analogue of vancomycin. The authors mimicked the natural system by anchoring the molecule to one of the bacterial 'peptide-carrier domains' involved in the biosynthesis. It could be that the enzymes responsible for forming the other two rings also work like this, so it will be well worthwhile examining their substrate specificity to see if that can be tweaked.

Joshua Finkelstein

General relativity

Galileo goes atomic

Phys. Rev. Lett. **93**, 240404 (2004)

Einstein reformulated Galileo's famous 'leaning tower' experiment as the weak equivalence principle, which proposes that the acceleration due to gravity is independent of a body's composition. Sebastian Fray *et al.* have tested this principle using atoms in place of Galileo's cannonballs.

If weak equivalence holds, two isotopes of rubidium (^{85}Rb and ^{87}Rb) should experience the same acceleration in free fall. To make accurate measurements of the atomic motions, the researchers used atom interferometry, which exploits the quantum wave-like nature of atoms. Such an approach has been used previously to compare the free fall of atoms with that of macroscopic objects. But comparing two different isotopes with each other enabled Fray *et al.* to investigate the effect of different quantum states on an atom's free fall.

In particular, the differences in relative orientation of electron and nuclear spin put the theoretical possibility of a spin-dependence of the gravitational force to the test. However, the authors find that, both for mass and for spin-state differences, the weak

Cell biology

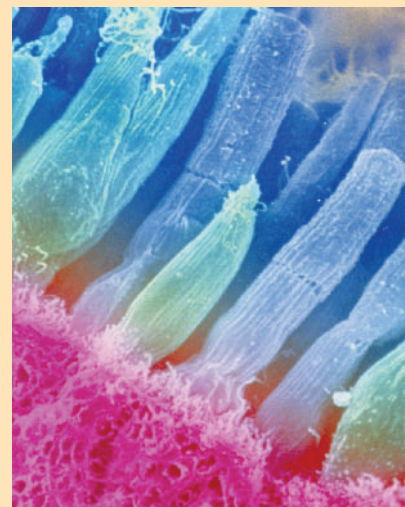
Retinal waste disposal

J. Exp. Med. doi:10.1084/jem.20041447

Rods and cones in the retina continuously slough off old fragments of their light-sensitive outer segments, following a 24-hour rhythm: rods (blue in picture) shed their aged tips every morning, whereas cones (blue-green) do so at night. These photoreceptor fragments are eaten up, or phagocytosed, by cells in the adjacent retinal pigment epithelium layer.

Emeline F. Nandrot *et al.* have found that disrupting this timed cycle of shedding and garbage disposal makes mice blind. They examined animals that had been genetically engineered to lack the integrin $\alpha_v\beta_3$ protein, which is used by pigment epithelial cells to take up the shed particles. In these mutant mice, the basal level of phagocytosis of photoreceptor fragments remained similar, but the epithelial cells lacked the early morning burst of activity that matches the rods' shedding. The mice slowly accumulated excess material and, after eight to ten months, went blind.

The authors suggest that the integrin triggers



signals in the epithelial cells that both control the rhythm of phagocytosis and ensure the timely and efficient digestion of consumed material. The animals' deterioration mirrors that seen in human retinal ageing and macular degeneration, and could serve as a model for these conditions.

Helen Pearson

equivalence principle seems to hold to an accuracy of about 1 part in 10^7 .

Philip Ball

Conservation

Shaky forest lifeline

Conserv. Biol. **18**, 1607–1616 (2004)

One hope for conservationists despairing at the current rate of extinctions has been that some forest species might be able to survive in the 'secondary' habitats that spring up in the aftermath of deforestation. But an analysis of Brazil's Atlantic forest ecosystems shows that, for birds at least, this lifeline does not seem to exist.

Destruction of primary forest is just as likely to result in the extinction of birds that have alternative habitats as of those that only use intact forest, report Grant M. Harris and Stuart L. Pimm. They measured original range size, current range size and abundance for 176 bird species, and estimated their extinction threat. The authors were looking for potential 'survivors' — species that, because they can make do in other habitats, would not die out if they lost their primary home. They found none.

Harris and Pimm suspect that the species spotted in secondary habitats are simply better at dispersing across the landscape, and that they presumably also make use of patches of unharmed primary forest. But the discovery that all bird species seem to be equally vulnerable to the complete removal

of primary forest underlines the importance of conserving these habitats — today, only around 10% of Brazil's Atlantic forest is still standing.

Michael Hopkin

Physics

Melting matters

J. Appl. Phys. **96**, 6127–6132 (2004)

The melting point of ice increases slightly when it is exposed to a strong magnetic field, Japanese scientists have found. This is surprising, given that water is diamagnetic and therefore has no intrinsic magnetic properties.

Hideaki Inaba and colleagues believe that the effect occurs because the applied magnetic field further stabilizes the network of hydrogen bonds that normally keeps the water molecules together. They showed that the change in melting point is proportional to the square of the magnetic field. For example, in a field of 6 tesla, the melting point of ice increased by 5.6 millikelvin.

Other researchers have noticed changes in the refractive index and the infrared spectrum of water under similar magnetic fields, again implicating hydrogen-bond stabilization. The hydrogen bond is mainly an electrostatic attraction between the slightly positively charged hydrogen atoms and slightly negatively charged oxygen atoms in a water molecule.

Inaba *et al.* suggest that the thermal motion of these charges in the magnetic field generates a force that damps down their movement, keeping the water molecules more tightly bound together.

Mark Peplow

Hydrothermal vent crabs feast on sea 'snow'

The riddle is solved of how vast crab populations can thrive in an extreme, toxic habitat.

The crab *Xenograpsus testudinatus* lives at enormously high densities around the sulphur-rich hydrothermal vents found in shallow waters off Taiwan, even though this acidic environment is low in nutrients. Here we show that these crabs swarm out of their crevices at slack water and feed on the vast numbers of zooplankton that are killed by the vents' sulphurous plumes, and that rain down like marine 'snow'. This opportunistic feeding behaviour explains how the crabs are able to survive in the adverse toxic environment of these shallow hydrothermal vents.

Investigation into the ecology of animals living in deep-water hydrothermal vents has been extensive^{1,2}, but much less is known about those living in shallow-water systems (at depths of less than 200 m)^{2–5}. Compared with the species-rich communities associated with deep-water vents^{1,2}, shallow-water vents are low in animal diversity and their ecosystems are simple — only two species of brachyuran crab, *X. novaeinsularis* and *X. testudinatus*, have been identified so far^{5–8}.

The shallow-water vents off Kueishan Island (121° 57' E, 24° 50' N; Fig. 1, and see supplementary information) in northeastern Taiwan are part of the Okinawa Arc⁹. In this unusual system, the vent discharges are highly acidic (pH 1.75–4.60) and sulphur-rich¹⁰ — with up to nine large smokers (2–6 m in height) at any one time spewing sulphurous plumes and bubbles of gas (mainly carbon dioxide, nitrogen, oxygen, sulphur dioxide and hydrogen sulphide), which emerge at temperatures of 65–116 °C.

High sulphide concentrations are normally important for the establishment of a chemolithotrophic food-web^{1,2} and the growth of sulphur-dependent bacterial mats¹¹. However, these food sources are not present in Kueishan, where the shallow-water vent discharges contain high concentrations of elemental sulphur (99.5% purity)¹⁰ and toxic volcanic gases (see supplementary information), so it is not surprising that the habitat is species-poor.

Almost nothing is known about the ecology of *Xenograpsus* crabs. The species *X. novaeinsularis* has been observed feeding on the ocean floor using its setae-tipped pincers⁷ but its diet is not known. At Kueishan, *X. testudinatus* congregates in large numbers in vent crevices^{6,8} (at an average density of 364 crabs per square metre; Fig. 2a, and see supplementary information), and the question arises as to how this ecosystem can



Figure 1 Kueishan Island. Sulphur particles are seen on the sea surface as a whitish area; the density of these particles varies with the number of chimneys under the sea and the prevailing wave and current conditions.

support such a substantial crab population. What are the crabs feeding on?

Our observations of the crabs at Kueishan show that *X. testudinatus* is feeding on zooplankton. During slack water, thousands of *Xenograpsus* crabs swarm out of the sulphur-rich crevices (Fig. 2b, and see movie in supplementary information) and begin to feed frantically on the sea floor over an area of a few square metres. This previously undescribed swarming behaviour is seen only during such slack-water periods. Dissection of *Xenograpsus* specimens revealed that their guts were full of zooplankton (mainly pelagic copepods). Studies of the mouthparts and

gastric mill of *Xenograpsus* confirm that it is a scavenger¹².

During optimal conditions, we witnessed a gradual precipitation of fine particulate matter and dead fish during slack water around the vents. Analysis of water samples collected at this time revealed a high concentration of inorganic particles and dead or narcotized zooplankton. During slack water, when currents are weak or absent, the vent plumes are directed vertically, instantly killing any organism in their path¹³ and causing it to drop straight down the water column. The mass of descending zooplankton has the appearance of falling snow.

As soon as the currents increase again and the deadly plume veers away, the bombardment by dead zooplankton ceases and the crabs return to their crevices. As tides in Taiwan are semidiurnal, crab feeding runs probably take place twice daily (although we have not confirmed this because diving at night is too dangerous).

The crabs time their foraging to coincide with these marine snowfalls in order to maximize their efficiency in harvesting the plankton kill. This opportunistic feeding behaviour by *Xenograpsus* crabs, which to our knowledge has not been seen in any other hydrothermal organism, represents a remarkable adaptation to their nutrient-poor environment.

M.-S. Jeng*, N. K. Ng†, P. K. L. Ng†

*Institute of Zoology, Academia Sinica, Nankang, Taipei, Taiwan 11529, Taiwan
e-mail: jengms@gate.sinica.edu.tw

†Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore

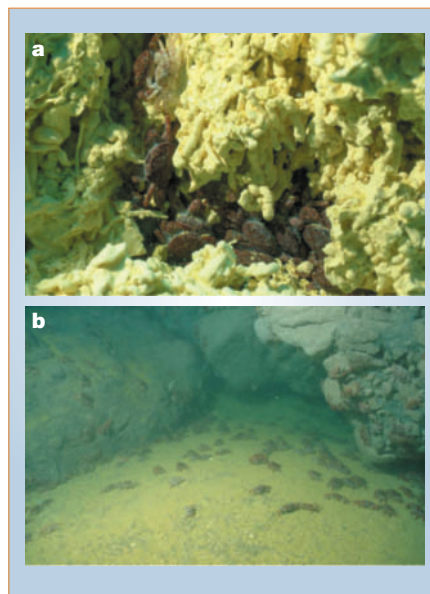


Figure 2 Crab feeding behaviour at shallow hydrothermal vents on the eastern side of Kueishan Island. **a**, Dense groups of *Xenograpsus* crabs in the crevices of sulphur-rich rubble at the base of a chimney. **b**, During a slack tide, the crabs swarm out to feed on the 'snowfall' of dead zooplankton that have been killed by toxic gases and sulphur particles, which billow from the vent.

1. Tunnicliffe, V. *Annu. Rev. Oceanogr. Mar. Biol.* **29**, 319–407 (1991).
2. Van Dover, C. L. *The Ecology of Deep-sea Hydrothermal Vents* (Princeton Univ. Press, New Jersey, 2000).
3. Dando, P. R., Hughes, J. A. & Thiermann, F. *Geol. Soc. Spec. Pub.* **87**, 303–317 (1995).
4. Southward, A. J. et al. *J. Mar. Biol. Assn. UK* **77**, 753–771 (1977).
5. Türkay, M. & Sakai, K. *Senckenberg. Mar.* **26**, 25–35 (1995).
6. Ng, N. K., Huang, J. F. & Ho, P. H. *Natl Taiwan Mus. Spec. Pub. Ser.* **10**, 191–199 (2000).
7. Takeda, M., Takeuchi, H. & Suganuma, H. *Nat. Environ. Sci. Res.* **6**, 59–64 (1993).
8. Jeng, M. S., Clark, P. F. & Ng, P. K. L. *J. Crust. Biol.* **24**, 188–212 (2004).
9. Liu, C. C. *Proc. Geol. Soc. China* **38**, 229–242 (1995).
10. Kuo, F. W. Thesis, Inst. Mar. Geol. Chem., Natl Sun Yat Sen Univ. (2001).
11. Childress, J. J. & Fisher, C. R. *Ocean. Mar. Biol. Annu. Rev.* **30**, 337–441 (1992).
12. Ng, N. K., Yang, S. & Ng, P. K. L. *Crustaceana* (submitted).
13. Kido, T. *J. Tokyo Univ. Fish.* **61**, 59–63 (1975).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Recent and episodic volcanic and glacial activity on Mars revealed by the High Resolution Stereo Camera

G. Neukum¹, R. Jaumann², H. Hoffmann², E. Hauber², J. W. Head³, A. T. Basilevsky^{1,4}, B. A. Ivanov^{1,5}, S. C. Werner¹, S. van Gasselt¹, J. B. Murray⁶, T. McCord⁷ & The HRSC Co-Investigator Team*

¹Institut fuer Geologische Wissenschaften, Freie Universitaet Berlin, Malteserstrasse 74-100, Bldg D, 12249 Berlin, Germany

²DLR-Institut fuer Planetenforschung, Rutherfordstrasse 2, 12489 Berlin, Germany

³Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

⁴Vernadsky Institute of Geochemistry and Analytical Chemistry, RAS, Kosygin Street 19, 119991 Moscow, Russia

⁵Institute of Dynamics of Geospheres, Leninskiy Prospect 38, 119334 Moscow, Russia

⁶Department of Earth Sciences, Open University, Milton Keynes MK7 6AA, UK

⁷Hawaii Institute of Geophysics and Planetology, University of Hawaii, 2525 Correa Road, Honolulu, Hawaii 96822, USA

*A list of all members of the HRSC Co-Investigator team and their affiliations appears at the end of the paper.

The large-area coverage at a resolution of 10–20 metres per pixel in colour and three dimensions with the High Resolution Stereo Camera Experiment on the European Space Agency Mars Express Mission has made it possible to study the time-stratigraphic relationships of volcanic and glacial structures in unprecedented detail and give insight into the geological evolution of Mars. Here we show that calderas on five major volcanoes on Mars have undergone repeated activation and resurfacing during the last 20 per cent of martian history, with phases of activity as young as two million years, suggesting that the volcanoes are potentially still active today. Glacial deposits at the base of the Olympus Mons escarpment show evidence for repeated phases of activity as recently as about four million years ago. Morphological evidence is found that snow and ice deposition on the Olympus construct at elevations of more than 7,000 metres led to episodes of glacial activity at this height. Even now, water ice protected by an insulating layer of dust may be present at high altitudes on Olympus Mons.

On board the European Space Agency (ESA) Mars Express Orbiter, a multiple line scanner instrument, the High Resolution Stereo Camera (HRSC), is acquiring high-resolution colour and stereo images of the surface of Mars¹. Resolution down to 10 m per pixel coupled with large areal extent (swaths typically 60–100 km wide and thousands of kilometres long) means that small details can be placed in a much broader context than was previously possible. Among the major objectives of the experiment is an assessment of the level of recent geological activity on Mars, particularly the type of volcanic and climate-related deposits that might indicate areas of hydrothermal activity and recent water exchange conducive to exobiological activity.

We have used the new HRSC images and their particular qualities in mapping out terrain types for the interpretation of morphological features and topographic relationships from the three-dimensional data and high-resolution imagery, including the Super Resolution Channel (SRC) data (resolution down to 2.5 m per pixel)¹. The high-resolution colour data were very useful for distinguishing different materials. The combined use of the HRSC data and nested Mars Orbiter Camera (MOC) or SRC imagery has proven to be extraordinarily helpful in the interpretation of the morphologies and processes that shaped the landforms now visible. Here we focused on the time-stratigraphic relationships and the sequence of events to understand the geological evolution of the martian areas investigated. Time sequences were obtained by determining the number of superimposed impact craters and deriving absolute ages.

This approach has become a powerful tool of planetary studies since the early 1970s, when frequencies of craters per unit area of lunar basaltic lavas were compared with the absolute ages of these lavas determined through isotopic dating of the returned samples

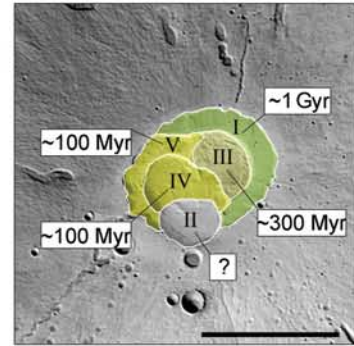
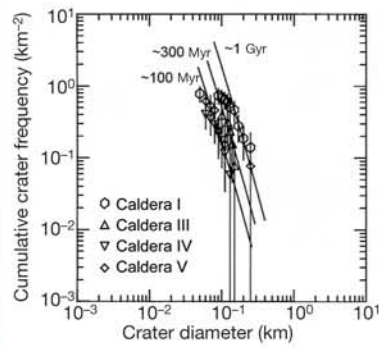
and have thus given us a reference scale for interplanetary comparison². These data, along with theoretical modelling and observational data of fluxes of crater-forming impactors for different parts of the Solar System, have made it possible to apply this method to different planets and satellites^{3–5}. For this study we used the recently updated cratering chronology model that combines the efforts of two major research groups in this area⁵.

The ages from crater counts are limited in accuracy almost exclusively by the statistical error³. Other error sources, such as undetected admixtures of secondary craters, volcanic or sublimation pits, are normally minor (<10% of the frequency of superposed craters)^{3,5}, provided the geological mapping of the areas and the counts are carried out by experienced observers. The statistical errors of individual data points in our counts are mostly <30% (one standard deviation, 1 σ). Because the whole distribution over a wider crater size range is used for fitting the theoretical size-frequency distribution to the measurements, the average statistical error of the data points over the ensemble of measurement is the proper measure for the uncertainty (which, in proportion to the number of data points, is much smaller), resulting in an average uncertainty of 20–30% in frequency. This translates into a 20–30% uncertainty in the absolute ages for ages younger than 3 gigayears (Gyr) and an uncertainty of only 100–200 million years (Myr) for ages older than 3 Gyr. Absolute ages may equally be affected by a possible systematic error of about a factor of two in the crater frequency for an assigned absolute age in the cratering chronology model used. This is due to an uncertainty in the underlying impact flux model used for Mars, relative to the lunar value^{4,5}.

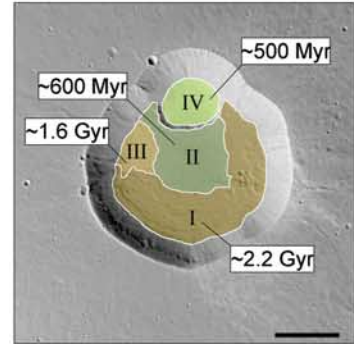
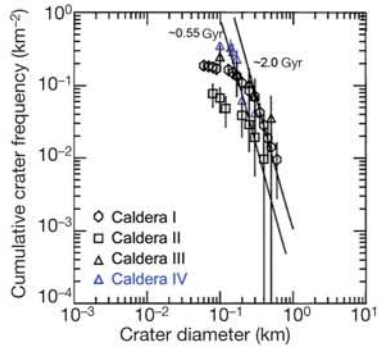
Time-stratigraphic relationships on martian volcanoes

The Tharsis region of Mars, a huge rise comprising almost 20% of

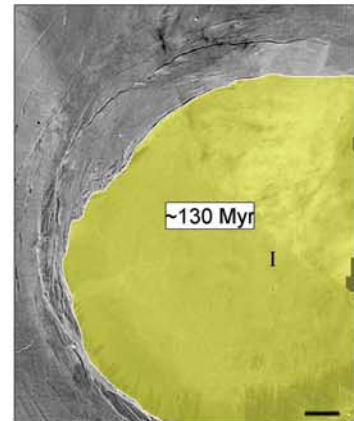
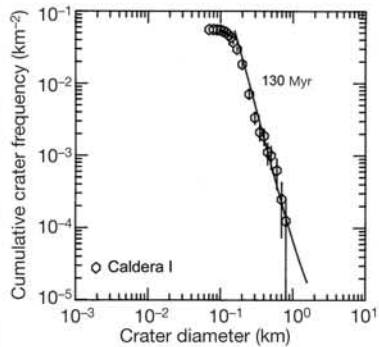
a Hecates Tholus



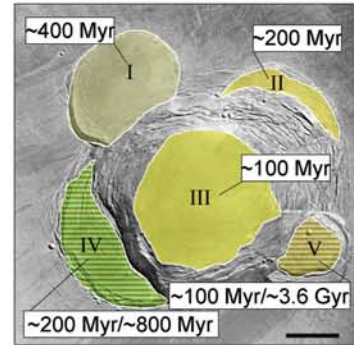
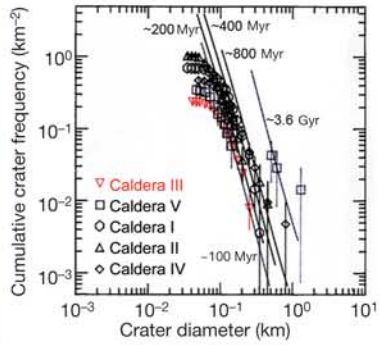
b Albor Tholus



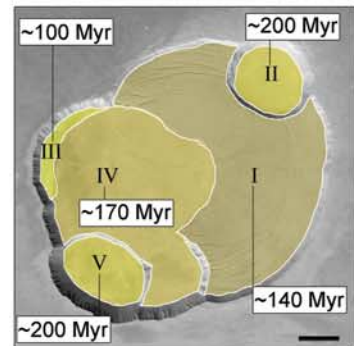
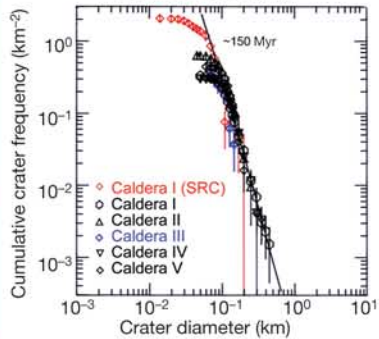
c Arsia Mons



d Ascraeus Mons



e Olympus Mons



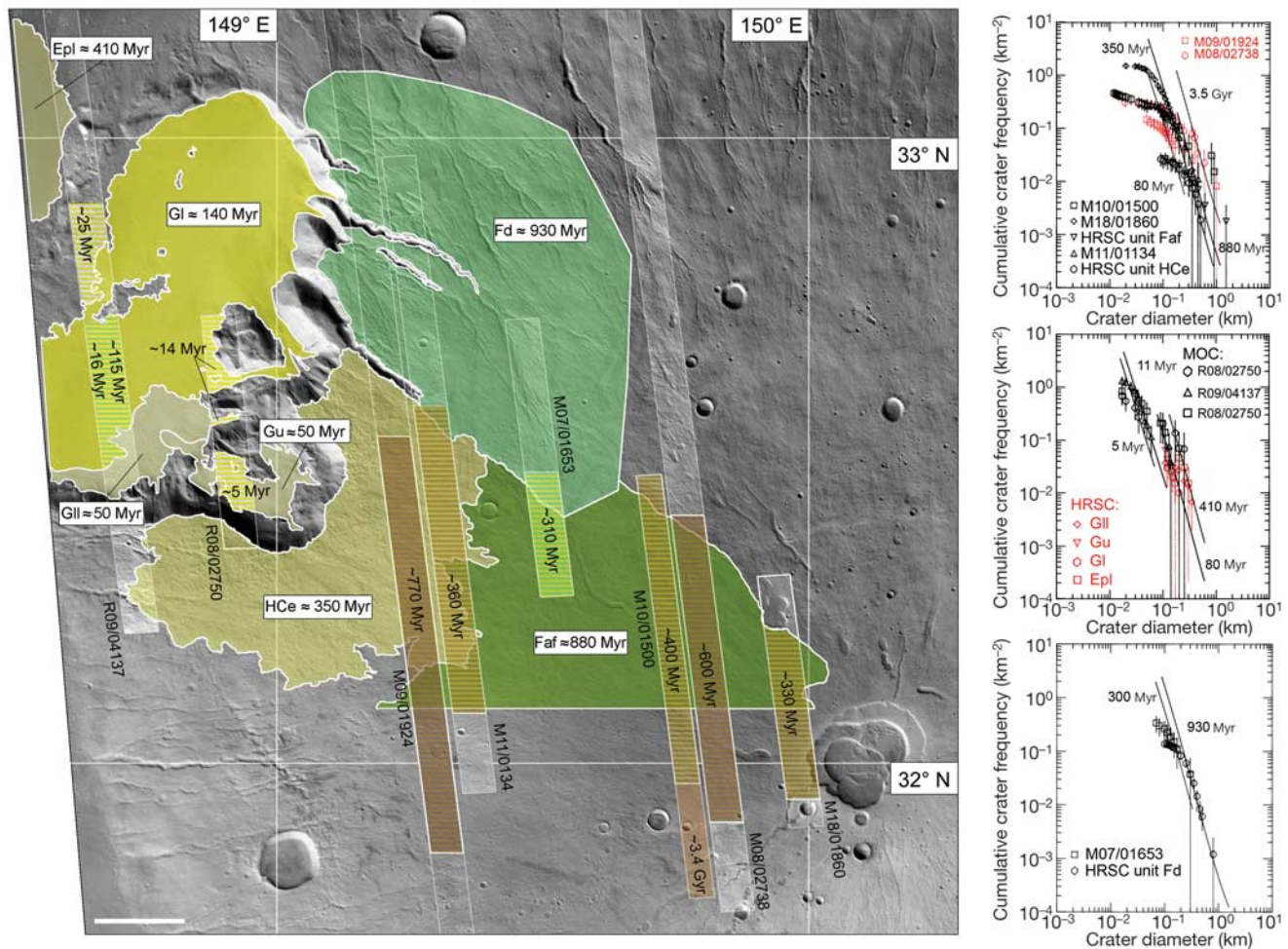


Figure 2 Hecates Tholus counting-area. Base map (HRSC image data and nested MOC data) (left panel), crater size-frequency data and derived absolute ages (three right

panels). Methodology as in Fig. 3. Scale bars, 10 km. North is at the top. Error bars, 1- σ error.

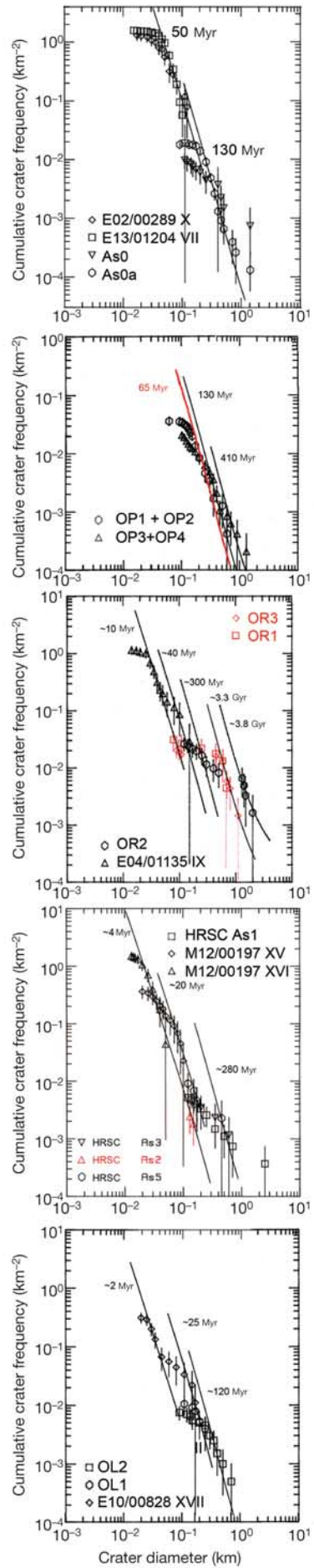
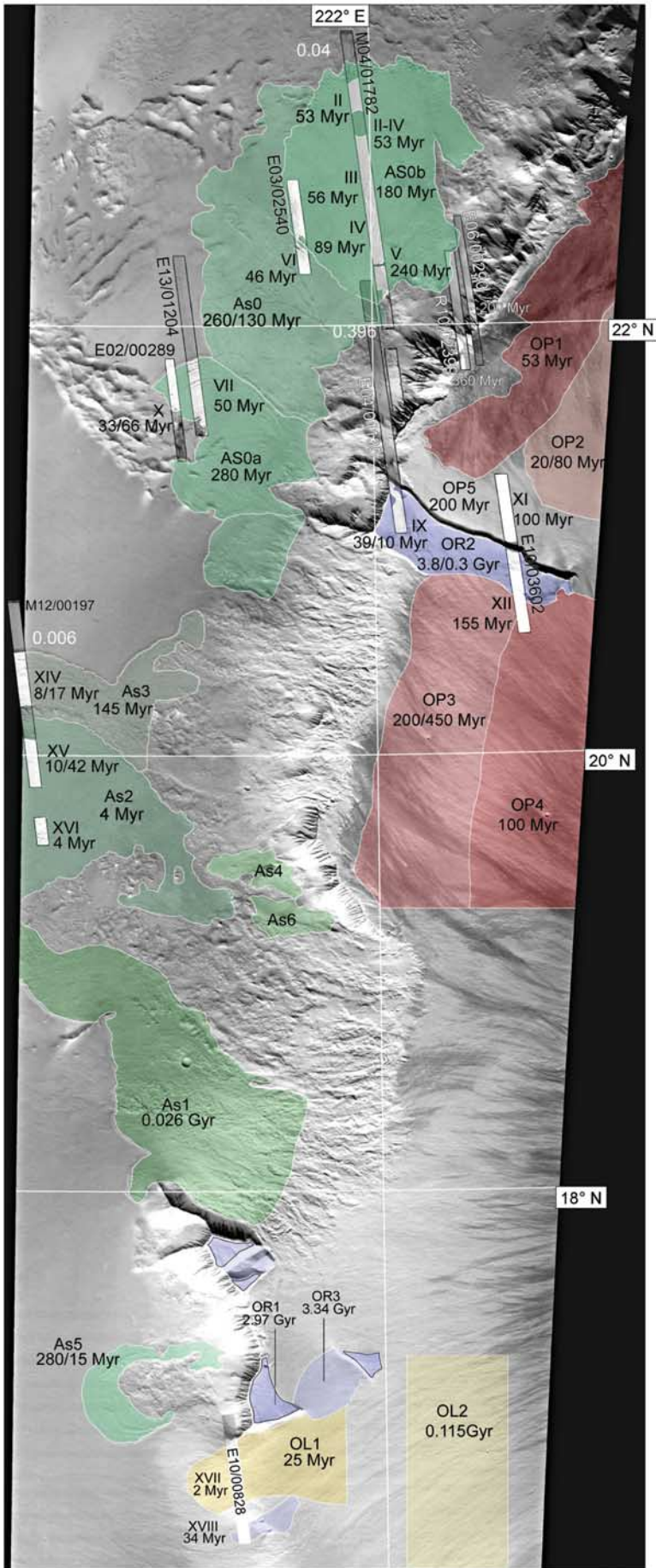
the planet, and the smaller Elysium rise have long been known to be the major focal points of volcanism over a significant portion of the history of Mars^{6,7}. However, the nature and extent of geologically very recent activity has been unknown. HRSC images acquired in the early phase of the mission (from January to July 2004) have provided broad high-resolution coverage that has cast new light on recent activity. Much of the relatively recent volcanic activity is focused on the summit calderas and flanking rift zones of the major shield volcanoes⁸ and therefore these areas were targeted early in the mission. Previous age determinations from crater counting have been limited either by poorer resolution or by the small areas imaged⁹.

New HRSC crater counts have confirmed the wide range of ages found earlier⁶ for the Tharsis and Elysium volcanism as a whole. For example, the ages of lavas on the flanks of Olympus Mons and Hecates Tholus range between 3,800 and 100 Myr, spanning 80% of the history of Mars (Supplementary Table 1). This extremely long-

lived history of volcanic activity is at least two orders of magnitude longer than the typical lifetime of large volcanoes on Earth, which are normally born, grow and become extinct in less than a million years. The very long activity of martian volcanoes implies correspondingly long lifetimes of 'hot spots' in the planet's interior. Hot spots on Earth are estimated to last for 100–200 Myr (ref. 10)—are martian hot spots different from terrestrial ones? Or does plate tectonics on our planet disturb the dynamics or limit our ability to trace back the total lifetimes of Earth's hot spots?

Evidence for geologically young volcanism is abundant in the calderas—depressions commonly found on the summits of volcanic edifices and representing collapses due to magma migration from subsurface reservoirs¹¹. The HRSC data show the ages of the floors of different collapse events within calderas on Olympus Mons, Ascaeus Mons, Arsia Mons, Albor Tholus and Hecates Tholus (Fig. 1 and Supplementary Table 1). Most show a complex history, with up to five different collapse episodes detectable, and widely differing ages of the resulting floor deposits (Fig. 1). On Ascaeus Mons, for example (Fig. 1d), the large central caldera floor is ~100 Myr old, and cuts adjacent caldera floors, which are interpreted to have formed ~200, ~400, and ~800 Myr ago, and perhaps earlier. The central caldera floor of Albor Tholus (Fig. 1b) is older than that of Ascaeus (~500 Myr), and is surrounded by portions of older caldera floors dating from ~600, ~1,600 and ~2,200 Myr ago. These findings are in agreement with theoretical analysis, which has suggested that subsurface magma reservoirs must cool and solidify

Figure 1 Investigated volcanic calderas. Base maps (left panels) and counting areas (right panels) of the investigated calderas derived from HRSC data, with crater size-frequency measurements and derived absolute ages (middle panels). Methodology as in Fig. 3. **a**, Hecates Tholus; **b**, Albor Tholus; **c**, Arsia Mons; **d**, Ascaeus Mons; **e**, Olympus Mons. Scale bars, 10 km. North is at the top. Error bars, 1- σ error,



between caldera collapse events, suggesting that magma supply to major shield volcanoes on Mars must have been episodic, rather than continuous¹².

Olympus Mons, however, is unusual by comparison with other calderas. At least five arcuate caldera wall segments can be identified (Fig. 1e), but instead of being spread over hundreds of millions of years in age span, as observed at Ascraeus, Albor and Hecates, the ages of the five Olympus Mons caldera floors cluster in the period 100–200 Myr ago. Theoretically, the ages should be older for high-lying caldera floors and younger for the caldera floors at lower elevation. The different ages are very close to each other within the error limits of ± 50 Myr of the age measurements. Thus, the formation of all calderas could have happened in a narrow time span around 150 Myr ago.

It is obvious, however, from the imagery that some of the caldera floors were slightly resurfaced by subsequent thin lava flows or tectonic processes¹³ such as horst and graben formation, accompanied by mass wasting processes. Therefore, it is more likely that the calderas formed and were modified subsequently within a period of several tens of millions of years (equivalent to the differences in ages around the average age of 150 Myr). If the theoretical predictions¹² are correct, this implies that separate magma reservoirs were forming, solidifying, and re-forming on timescales averaging perhaps 20 Myr apart. Furthermore, the summit of Arsia Mons is dominated by a single huge caldera whose floor is dated at ~ 130 Myr ago—falling within the time span represented by the five Olympus Mons ages (~ 100 –200 Myr ago).

These ages confirm some earlier measurements^{5,9} on the basis of MOC images in small areas of the calderas of Olympus Mons and Arsia Mons and indicate that the summits of these edifices were very active in essentially the geological present, the last 2–4% of Mars history. These ages provide supporting evidence for the repetitive and episodic nature of caldera formation, and thus magma supply, to the major shield volcanoes on Tharsis and Elysium (compare also with Hecates Tholus in Fig. 2). It is also an interesting coincidence that the summits of four of five of these edifices were very active 100–200 Myr ago, the period in which the crystallization ages of one of two major groups of martian basaltic meteorites fall¹⁴. This does not necessarily imply a genetic relationship but is a clue suggesting probably widespread volcanic activity on Mars at that time, generating large surface units that the basaltic meteorites may have come from.

It has long been known that the flanking rift zones of the Tharsis

Montes and lava flows cascading over the Olympus Mons scarp postdate much of the central edifice-building activity¹⁵. The new HRSC data permit more precise dating of the duration of activity in these regions. For example, on the lower flanks of Olympus Mons (Fig. 3) are observed flows for which crater size-frequency and age characteristics are interpreted to be representative of activity at ~ 115 Myr ago, ~ 25 Myr ago, and with HRSC and MOC data combined, as recently as 2.4 Myr ago.

Hydrothermal, fluvial, and glacial activity

Further evidence of very recent and episodic geological activity on Mars has been obtained by HRSC in the form of images and ages of several deposits related to recent climate change. We know that water in the form of ice exists at the polar caps and in the cryosphere of Mars¹⁶, and perhaps locally on the surface¹⁷. Only recently, however, has it become clear that the extreme variability of the obliquity of the spin axis of Mars and orbit eccentricity¹⁸ can cause significant mobilization of polar volatiles and their redeposition equatorward^{19,20}. Of particular interest are the types of deposits that are interpreted to represent the accumulation of water ice in non-polar regions, because these are very sensitive environmental indicators and have important implications for possible life and future automated and human exploration.

For these reasons, early targets for the HRSC instrument were parts of the Elysium region, and the western scarp of the Olympus Mons volcano. On the basis of Viking imagery, channels on the flanks of the Hecates shield have already been detected and interpreted as having been produced by running water^{21,22}. We have been able to study the Hecates shield at a resolution of 26 m per pixel in great detail and to determine the ages of some areas using crater-statistics methods (Fig. 2). MOC data were also used, as indicated in Fig. 2. The data show a wide age range over which volcanic activity and related mobilization of water (probably released hydrothermally or partly released through melting of snow caps by volcanically induced heating from underground) with subsequent glacial activity occurred. Here, we present only the gross time-stratigraphic relationships of the development in different areas on the volcano, starting more than 3.4 Gyr ago and shaping the volcano through different episodes of activity (for example, ~ 900 , 400 and 50 Myr ago) until very recent times of about 5 Myr ago. Fluvial and glacial activity can be recognized close to or in the depressions at the northwestern base of the volcano. In southern Elysium, to the southwest of Athabasca Valles, surface features have been observed on the HRSC imagery that look similar to pack-ice on Earth. The age of these deposits is only 5 Myr—in the same range as some of the glacial deposits on Olympus Mons and Hecates Tholus. Details of our findings are supplied elsewhere^{23,24}.

The other outstanding early target, the Olympus Mons volcano (Figs 3 and 4), is a site known to be characterized by lobate deposits thought to be of glacial origin¹⁷ and recently shown on the basis of the NASA Mars Global Surveyor (MGS) and Odyssey mission data to be a series of lobate rock-covered piedmont glaciers²⁵. These deposits are well illustrated in the new HRSC data (Fig. 4b). Therefore, it is now possible to determine more specific ages for them (Fig. 3 and Supplementary Table 1): 130 to 280 Myr for the major lobes, with some subunits in the 20–60 Myr range and locally as young as 4 Myr. These data indicate that the lobate deposits represent several phases of formation, most probably representing periods when significant snow and ice accumulation (possibly accompanied by hydrothermal mobilization of water that flowed down over the edge of the shield, entraining large amounts of non-icy surface material and then freezing) at the Olympus Mons scarp caused mobilization and flow of debris-covered piedmont glaciers into the surrounding low-lying regions.

In several places along the scarp, small linear tongue-like deposits can be seen to emerge from within the apparent accumulation zones of the larger lobate deposits (Fig. 4c). These are interpreted to be

Figure 3 Olympus Mons western scarp areas. HRSC-image base map with nested MOC data and depiction of the counting areas (left panel) and the resulting ages of the western near-escarpment area of the Olympus Mons volcanic shield, the 7-km-high escarpment and the adjacent plains area to the west with remnants of glacial features (five right panels). The counts show different episodes of resurfacing with erosion of craters and subsequent re-cratering. These episodes and processes are reflected in the different steepnesses of the distributions on the log–log plots, giving a kinked appearance. The flat parts show erosional effects; the steep parts show the re-cratering after the erosional episodes. The martian impact-crater size-frequency distribution^{12,13,14} has been fitted to the individual segments of the distribution, giving individual crater-frequency values for the different episodes; by application of the Hartmann–Neukum chronology¹⁴ individual absolute ages can be extracted. In this way it is possible to extract the evolutionary history of the area under investigation in detail. Here the fits to the crater frequencies partly have the character of average isochrons for a group of counts yielding similar numbers. Individual ages may be slightly different and are precisely given in Supplementary Table 1. The errors of the ages are usually around 20–30% for ages younger than 3 Gyr (only 100–200 Myr for older ages) owing to the statistical limitations. The error bars given represent a 1- σ error. In the same way, all ages of less than 2 Gyr may be affected by a possible systematic error of about a factor of two in the cratering chronology model¹⁴ used. North is at the top.

debris-covered alpine glaciers²⁶ that were previously undetected. Remarkably, the ages of these tongue-like deposits are so young that they cannot be reliably dated because of the lack of craters on them. These smaller tongue-like deposits are interpreted to have formed when conditions were sufficient to cause ice accumulation and local flow, but obviously did not form over the entire extended duration

represented by the underlying more widespread deposits. These very young tongue-like deposits are characterized by depressions at the base of the scarp and thus apparently lack an active accumulation zone, suggesting that they are relict features that are no longer forming today.

HRSC and local MOC data reveal evidence for sedimentary

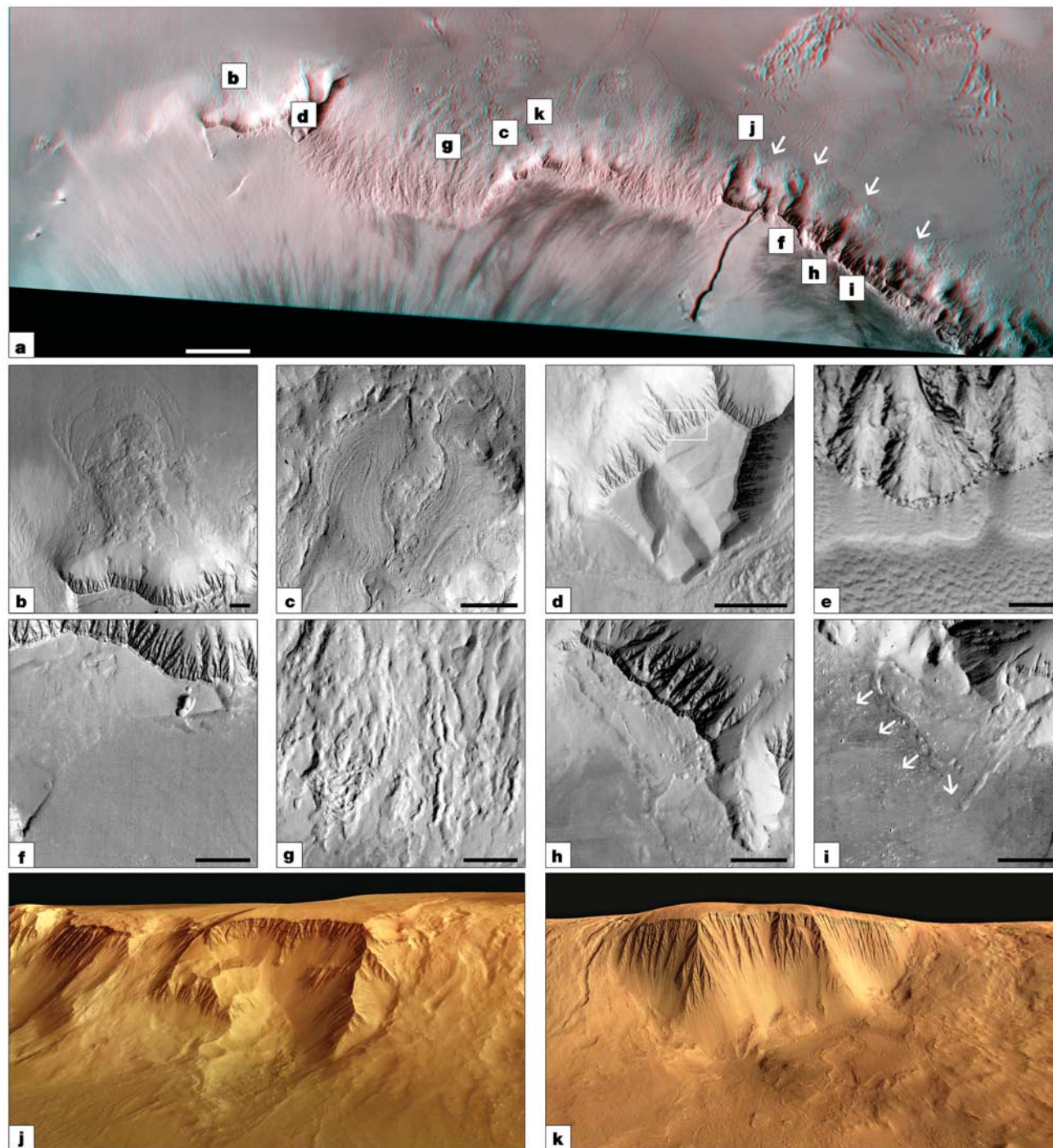


Figure 4 Ice-dust deposits and glaciers on Olympus Mons. **a**, Colour anaglyph for three-dimensional view of the region (scale bar, 30 km); labels b–k indicate designated areas shown on panels b–k (scale bar, 2 km). **b**, Glacier-type lobate flow. **c**, Tongue-like flows. **d**, Mesa at the scarp edge; small white square shows area of panel **e**. **e**, Part of MOC image E05-02498 showing fine layering in the upper part of the mesa material. **f**, Non-impact pits on mesa (upper right) and lava flows entering the mesa (centre left). **g**, Collapse-type depression and channel. **h**, The edge ridge with layers and the upper part of

U-shaped valley cut into the ridge and summit plateau. **i**, Glacier-type feature rimmed with small ridges (see arrows). **j**, **k**, Perspective views (from the west) of the western scarp of Olympus Mons showing steep ravined slopes and more gentle slopes with chaos-type depressions in their upper parts, as well as fluvial-type channels and glacier-like flows. The perspective view has been produced through a combination of HRSC nadir and colour image data and the DTM derived from the HRSC stereo data. North is to the right.

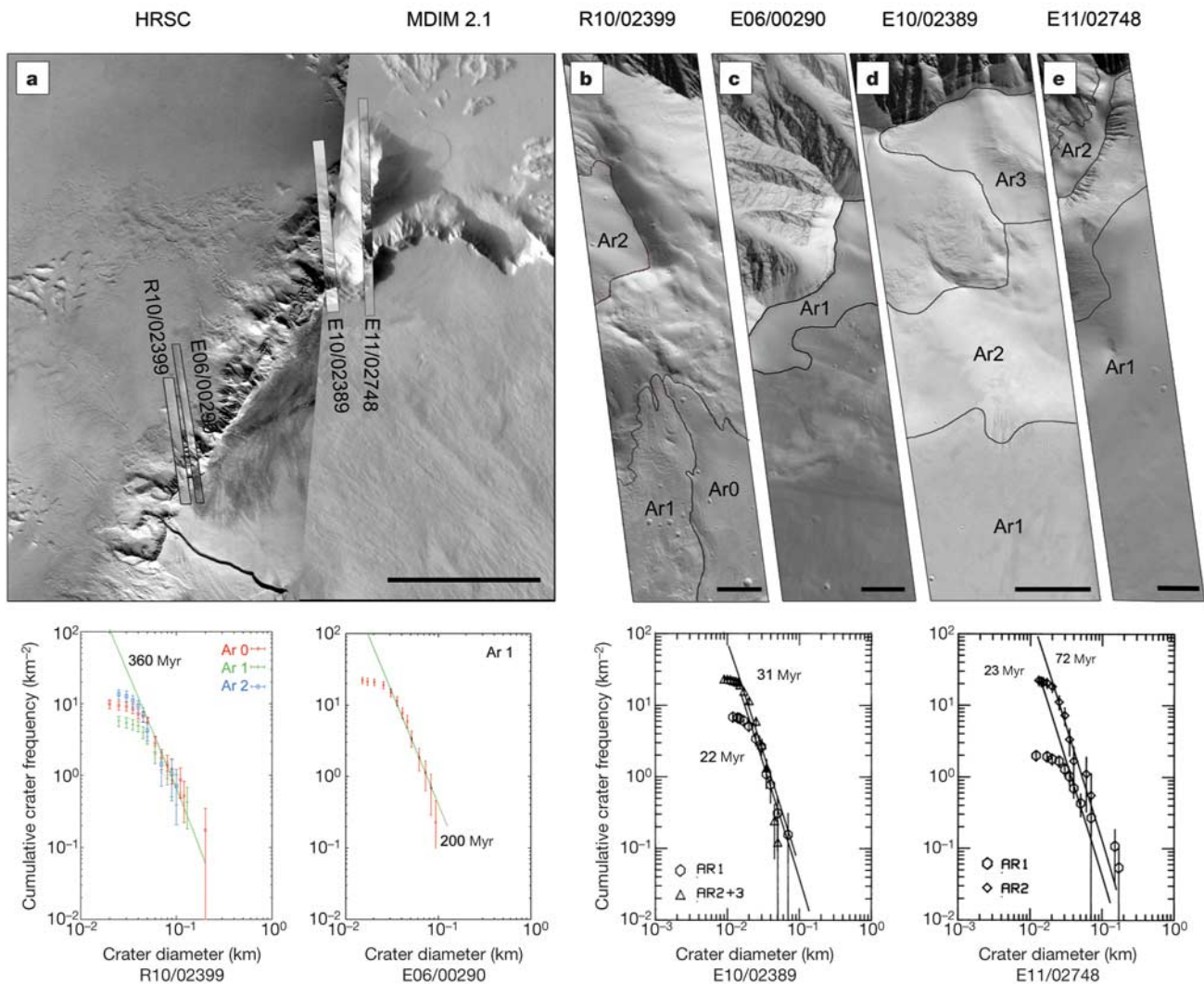


Figure 5 Base map (panels **a–e**) and crater statistics (four panels at bottom) of the areas on the northwestern part of Olympus Mons investigated for possible ice–dust coverage and age relationships. **a**, Mosaic of HRSC data (left) and Viking imagery (right) with the location of the MOC images used for morphological study and age determination; scale bar, 50 km. **b–e**, MOC images (scale bar, 1 km) with counting areas on flat terrain on the escarpment ridge with its ice–dust cover. **b**, Areas of measurement on MOC image R10/02399 and resulting ages derived from superposed crater frequencies. The ages are roughly the same for all three counting areas on the basis of the fit for intermediate-sized to large craters: 360 Myr with an uncertainty of about 70 Myr. Minor erosion of craters can be recognized in the subtle flattening of the measured distribution in comparison with the

fitted theoretical production function. The bending-over characteristics at the small-crater end here and in the other measurements in **c–e** stem from loss of craters at the resolution limit of the images. **c**, Areas of measurement on MOC image E06/00290 and resulting age of 200 ± 40 Myr. **d**, Areas of measurement on MOC image E10/02389 and resulting ages of 22 ± 4 Myr and 31 ± 6 Myr. The ages of the individual counting areas AR2 and AR3 have been determined separately with the same results (31 Myr). The AR1 measurements show that some erosion of the ice–dust cover seems to have taken place. **e**, Areas of measurement on MOC image E11/02748 and resulting ages of 23 ± 5 Myr and 72 ± 15 Myr. These ages correlate well with the ages (20 and 80 Myr) measured on adjacent lava flows (OP2 of the base map in Fig. 3). North is at the top. Error bars, 1- σ error.

deposits 50–100 m thick on the rim of the Olympus western scarp, sometimes in the form of mesas (Fig. 4d, e). These optically bright sediments show fine layering obviously different from that which could potentially be formed by accumulation of the lava flows typical of this volcano. We interpret these deposits to represent the remnants of accumulations of dust and ice high on the Olympus construct in previous geological periods. We believe that the presence on some mesas of rimless and deep sinkhole-type depressions (Fig. 4f, upper right) confirms this interpretation. In some places we also see that lava flowing upon these deposits causes their collapse (Fig. 4f, centre left).

Age measurements (Figs 3 and 4) show that the layers formed more than 300 Myr ago, possibly followed by further episodic activity or subsequent major erosional activity until about 65 Myr

ago, as seen on the escarpment ridge. Some data for the mesas, from the frequency of some large surviving craters that belong to the underlying substrate, yield age estimates that exceed 3 Gyr. That is close to earlier estimates of the age of the Olympus Aureole, interpreted to have formed by 3.4-Gyr-old gravitation slides of material from the western shield of the Olympus construct^{27,28}. This is an indication that the shield of Olympus Mons had already reached heights in excess of 7,000 m very early in its history. The eruption rates must have been very high at the beginning of the growth of the volcano.

Within the broad segments of the upper part of the scarp we observe depressions (Fig. 4g) that are morphologically similar to collapse depressions seen at the source areas of the outflow channels of Mars. The latter are believed to form as a result of release of large

amounts of water from below the surface, for example, the melting of ground ice by magma intrusion^{28,29}. Water (subsequently freezing to ice) seems to have been mobilized out of the ground hydrothermally. The collapse-type depressions of the Olympus scarp are also often accompanied by fluvial-type channels, although of relatively small size (Fig. 4h). We interpret these as having been formed by melting ice or ice-and-dust deposits via interaction with lavas.

Glaciers on the shield of Olympus Mons

The edge of the northern part of the scarp is rimmed by a ridge also covered with fine-layered deposits (Fig. 4j, k and Fig. 5a–e). The HRSC-based digital terrain model (DTM) model shows that the ridge stands 400–700 m above the surface adjacent to the east. We interpret the deposits on the ridge as a layer of ice and dust now protected from sublimation by a lag of dust. This ‘ice cap’ on the ridge formed about 400 Myr ago or earlier, as determined by measuring the superposed crater frequencies as given in Fig. 5. We identify different episodes of subsequent resurfacing of the ice cap at times of 200, 70 and 20–30 Myr ago. We have identified several valleys that are transverse to the ridge and cut into it. In a few cases, the valleys cut back into the summit plateau. The valleys have a U-shaped cross profile and end at the scarp foot with hummocky piles (arrows on Fig. 4a) often having lobate outlines. These are interpreted as debris-covered piedmont glaciers. These characteristics of the valleys suggest that glaciers ploughed them and that the source areas of those glaciers were high on the summit plateau. Moreover, in the potential source area of one of these valleys we see a lobate landform whose orientation suggests glacier movement to the east, towards the volcano centre (Fig. 4i). We estimate the age of this landform—now covered by dust, as are almost all other landforms of the region—to be 200–300 Myr (Figs 3 and 4j). We see evidence of sublimational or eolian degradation significantly erasing the cratering record, which implies that the lobate landform has not been active recently.

The colour data have been examined for effects of the presence of water ice, such as brightening of the surface and possibly flattening of the spectral slope. Colour ratio measurements were done for 16 major terrain types. Brightness in the four HRSC colour channels (centre wavelengths 440, 530, 750 and 970 nm) was calculated by applying the scaling factors from the ground calibration and additional factors derived from comparison with telescopic spectral data³⁰ and cross-calibration to data from the OMEGA experiment³¹. The measurements showed that the brightness values for each colour channel vary only within a factor of 1.5. All the measured ratios are close to those of ‘bright regions of Mars’³² indicating that all the studied terrains are dust-covered. If there is ice, then it appears to be covered and protected from sublimation by a sublimation lag of dust on top of the ice or ice–dust deposit.

The accumulation of ice necessary for developing the high-altitude glaciers and the dust–ice deposits on the mesas and the ridge could have happened at times of high inclination of the planet’s spin axis³³. Experimental data³⁴ and calculations³⁵ show that under current climate conditions sun-illuminated water ice (snow) sublimates quickly. But even a thin dust cover shifts the vapour pressure in the pore space to close to the saturation pressure and drastically decreases the sublimation rate of the buried ice³⁶. The buried ice (more probably an ice and dust mixture) may be practically stable (slowly sublimating) over millions of years, depending on the diffusive resistance of the upper dust cover against the vapour outflow to the dry atmosphere. Therefore, the layered deposits on the mesas and on the escarpment ridge, and the high-altitude lobate landform may still contain ice. This suggestion may be tested in the future by the MARSIS sounding radar³⁷ and OMEGA^{31,38} experiments on board Mars Express.

Conclusions

The new ages from the HRSC data: (1) confirm the very wide age range (billions of years) over which the Tharsis and Elysium regions were volcanically active; (2) reveal that summit caldera activity was periodic and often consistent with theoretical predictions of magma reservoir cooling and regeneration behaviour; (3) show that the most recent summit caldera activity on the Tharsis volcanoes was clustered ~100–200 Myr ago, practically coinciding with radiometric ages of several martian meteorites; (4) reveal that some of the youngest volcanism on the Tharsis edifices appears to be as young as several million years, thus suggesting that these volcanoes could well erupt in the future; (5) yield evidence for former hydrothermal mobilization of water at the western edge of the Olympus Mons volcano shield and probably on the Hecates Tholus volcano with subsequent development of glaciers; (6) reveal evidence for very young glaciations in the tropical regions of Mars; (7) reveal evidence for deposition of dust and ice and episodes of glaciation high on the Olympus Mons construct; and (8) suggest that water ice may now be present at high altitudes on the edge of the Olympus western scarp. □

Received 3 September; accepted 30 November 2004; doi:10.1038/nature03231.

- Neukum, G., Jaumann, R. & the HRSC Co-Investigator and Experiment Team. *HRSC—The High Resolution Stereo Camera of Mars Express* 17–35 (European Space Agency Special Publication ESA SP-1240, 2004).
- Malin, M. C. & Edgett, K. S. Mars Global Surveyor Mars Orbiter Camera: Interplanetary cruise through primary mission. *J. Geophys. Res.* **106**, 23429–23570 (2001).
- Neukum, G., Ivanov, B. A. & Hartmann, W. K. Cratering record in the inner Solar System in relation to the lunar reference system. *Space Sci. Rev.* **96**, 55–86 (2001).
- Ivanov, B. A. Mars/Moon cratering rate ratio estimates. *Space Sci. Rev.* **96**, 87–104 (2001).
- Hartmann, W. K. & Neukum, G. Cratering chronology and the evolution of Mars. *Space Sci. Rev.* **96**, 165–194 (2001).
- Neukum, G. & Hiller, K. Martian ages. *J. Geophys. Res.* **86**, 3097–3121 (1981).
- Tanaka, K. The stratigraphy of Mars. *J. Geophys. Res.* **91**, E139–E158 (1986).
- Greeley, R. & Spudis, P. D. Volcanism on Mars. *Rev. Geophys. Space Phys.* **19**, 13–41 (1981).
- Hartmann, W. K. *et al.* Recent volcanism on Mars from crater counts. *Nature* **397**, 586–589 (1999).
- Steinberger, B. Plumes in a convecting mantle: Models and observations for individual hotspots. *J. Geophys. Res.* **105**, 11127–11152 (2000).
- Crumpler, L. C., Head, J. W. & Aubele, J. C. Calderas on Mars: Characteristics, structure and associated flank deformation. In *Volcano Instability on the Earth and Planets* (eds McGuire, W. J., Jones, A. P. & Neubeck, J.) *Geol. Soc. Spec. Pub.* **110**, 307–348 (1996).
- Wilson, L., Scott, E. D. & Head, J. W. Evidence for episodicity in the magma supply to the large Tharsis volcanoes. *J. Geophys. Res.* **106**, 1423–1433 (2001).
- Hauber, E. *et al.* The calderas on Mars—magmatic and tectonic characteristics as revealed through images of the High Resolution Stereo Camera (HRSC) on Mars Express. *Eur. Geophys. Union 1st Gen. Assembly* (abstr.) EGU04-A-07922 (2004).
- Nyquist, L. E. *et al.* Ages and geologic histories of Martian meteorites. *Space Sci. Rev.* **96**, 105–164 (2001).
- Crumpler, L. S. & Aubele, J. C. Structural evolution of Arsia Mons, Pavonis Mons and Ascraeus Mons: Tharsis region of Mars. *Icarus* **34**, 496–511 (1978).
- Carr, M. *Water on Mars* 229 (Oxford Univ. Press, New York, 1996).
- Lucchitta, B. K. Mars and Earth: Comparison of cold climate features. *Icarus* **45**, 264–303 (1981).
- Laskar, J. *et al.* Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* **170**, 343–364 (2004).
- Richardson, M. I. & Wilson, J. R. Investigation of the nature and stability of the Martian seasonal water cycle with a general circulation model. *J. Geophys. Res.* **107** doi:10.1029/2001JE001536 (2002).
- Haberle, R. M., Murphy, J. R. & Schaeffer, J. Orbital change experiments with a Mars general circulation model. *Icarus* **162**, 66–89 (2003).
- Mouginis-Mark, P. J., Wilson, L. & Head, J. W. Explosive volcanism on Hecates Tholus, Mars: Investigation of eruption conditions. *J. Geophys. Res.* **87**, 9890–9904 (1982).
- Gulick, V. C. & Baker, V. R. Origin and evolution of valleys on Martian volcanoes. *J. Geophys. Res.* **95**, 14325–14344 (1990).
- Hauber, E. *et al.* Discovery of a flank caldera and very young and glacial activity at Hecates Tholus, Mars, in Mars Express HRSC images. *Nature* (submitted).
- Murray, J. B. Evidence from the Mars Express High Resolution Stereo Camera for a frozen sea close to Mars’ equator. *Nature* (2004) (submitted).
- Milkovich, S. M. & Head, J. W. Olympus Mons fan-shaped deposit morphology: Evidence for debris glaciers. *6th Int. Mars Conf.* (abstr.) 3149 (2003).
- Head, J. W. *et al.* Recent ice ages on Mars. *Nature* **426**, 792–802 (2003).
- Harris, S. A. The aureole of Olympus Mons. *J. Geophys. Res.* **82**, 3099–3107 (1977).
- Lopes, R., Hiller, K., Neukum, G. & Guest, J. E. Further evidence of the Olympus Mons Aureole. *J. Geophys. Res.* **87**, 9917–9928 (1982).
- Baker, V. R. *The Channels of Mars* (Austin Univ. of Texas Press, Austin, 1982).
- McCord, T. *et al.* The color capabilities of the Mars Express High Resolution Stereo Camera. *Eur. Geophys. Union 1st Gen. Assembly* (abstr.) EGU04-A-06358 (2004).
- Bibring, J.-P. *et al.* OMEGA: Observatoire pour la Minéralogie, l’Eau, les Glaces et l’Activité. 37–49 (European Space Agency Special Publication ESA SP-1240, 2004).
- Banin, A., Clark, B. C. & Waenke, H. in *Mars* (eds Kiefer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 594–625 (Univ. Arizona Press, Tucson, 1992).

33. Head, J. W. *et al.* Recent mid-latitude glaciation on Mars: Evidence for snow and ice accumulation and flow in Mars Express HRSC data. *Nature* (in the press).
34. Kossacki, K. J., Komle, N. I., Leliwa-Kopystynski, J. & Kargel, G. Laboratory investigation of the evolution of cometary analogs: results and interpretation. *Icarus* **128**, 127–144 (1997).
35. Kossacki, K. J., Markiewicz, W. J., Skorov, Y. V. & Komle, N. I. Sublimation coefficient of water ice under simulated cometary-like conditions. *Planet. Space Sci.* **47**, 1521–1530 (1999).
36. Skorov, Y. V., Markiewicz, W. J., Basilevsky, A. T. & Keller, H. U. Stability of water ice under a porous nonvolatile layer: implications to the south polar layered deposits of Mars. *Planet. Space Sci.* **49**, 59–63 (2001).
37. Picardi, G., *et al.* MARSIS: Mars Advanced Radar for Subsurface and Ionosphere Sounding. 51–69 (European Space Agency Special Publication ESA SP-1240, 2004).
38. Bibring, J. P. *et al.* Perennial water ice identified in the south polar cap of Mars. *Nature* **428**, 627–630 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank U. Wolf for help with the crater size-frequency distribution measurements and age evaluation, as well as W. Zuschneid and O. Fabel for their technical assistance and T. Denk for support in the colour data reduction effort. We also thank the ESTEC and ESOC staff, and the DLR Experiment and Operations team, especially T. Roatsch, K.-D. Matz, V. Mertens, J. Flohrer, F. Scholten, and K. Gwinner. J. Korteniemi, M. Aittola, P. Kostama and D. Williams helped with the image planning. This work forms part of the HRSC Experiment of the ESA Mars Express Mission and has been supported by the German Space Agency (DLR) on behalf of the German Federal Ministry of Education and Research (BMBF). Part of the data evaluation is supported by a grant from the German Science Foundation (DFG) within the scope of the priority programme ‘Mars and the Terrestrial Planets’.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G. N. (gneukum@zedat.fu-berlin.de).

The HRSC Co-Investigator Team (Authors are arranged in alphabetical order) J. Alibert¹, A. T. Basilevsky², G. Bellucci³, J.-P. Bibring⁴, M. Buchroithner⁵, M. H. Carr⁶, E. Dorrer⁷, T. C. Duxbury⁸, H. Ebner⁹, B. H. Foing¹⁰, R. Greeley¹¹, E. Hauber¹², J. W. Head III¹³, C. Heipke¹⁴, H. Hoffmann¹², A. Inada^{15,18}, W.-H. Ip¹⁶, B. A. Ivanov¹⁷, R. Jaumann¹², H. U. Keller¹⁸, R. Kirk¹⁹, K. Kraus²⁰, P. Kronberg²¹, R. Kuzmin², Y. Langevin⁴, K. Lumme²², W. Markiewicz¹⁸, P. Masson²³, H. Mayer⁷, T. B. McCord²⁴, J.-P. Muller²⁵, J. B. Murray²⁶, F. M. Neubauer²⁷, G. Neukum (Principal Investigator)²⁸, J. Oberst¹², G. G. Ori²⁹, M. Pätzold²⁷, P. Pinet³⁰, R. Pischel¹², F. Poulet⁴, J. Raitala³¹, G. Schwarz³², T. Spohn¹² & S. W. Squyres³³

Affiliations for authors: 1, TU Berlin, D-10623 Germany; 2, Vernadsky Institute-RAS, Moscow 117975, Russia; 3, CNR/IFSI, 00133 Rome, Italy; 4, CNRS/IAS, 91405 Orsay cedex, France; 5, TU Dresden, D-01062 Germany; 6, USGS, MS 975 Menlo Park, California 94025, USA; 7, Universität der Bundeswehr München, D-85577 Neubiberg, Germany; 8, JPL, Pasadena, California 91109, USA; 9, TU München, D-80333 Germany; 10, ESTEC/SCI-SR, Postbus 299, NL-2200 AG Noordwijk, The Netherlands; 11, Arizona State University, Box 871404, Tempe, Arizona 85287-1404, USA; 12, DLR, D-12489 Berlin, Germany; 13, Brown University, Box 1846, Providence, Rhode Island 02912, USA; 14, Universität Hannover, D-30167 Germany; 15, Kobe University, 657-8501 Kobe, Japan; 16, National Central University (NCU), Chung-Li 320, Taiwan; 17, IDG-RAS, Moscow 117979, Russia; 18, MPAE, Postfach 20, D-37191 Katlenburg-Lindau, Germany; 19, USGS, Flagstaff, Arizona 86001, USA; 20, TU Wien, A-1040 Wien, Austria; 21, TU Clausthal, D-38678 Clausthal, Germany; 22, University of Helsinki, PO Box 14, SF-00014 Helsinki, Finland; 23, OrsayTerre, F-91405 Orsay Cedex, France; 24, PSI-Nw, Winthrop, Washington 98862, USA; 25, University College London WC1E 6BT, UK; 26, The Open University, Milton Keynes MK7 6AA, Buckinghamshire, UK; 27, Universität Köln, D-50923 Germany; 28, FU Berlin, D-12249, Germany; 29, IRSPS, I-65127 Pescara, Italy; 30, Observatoire de Midi-Pyrénées, F-31400 Toulouse, France; 31, University of Oulu, FIN-90401 Finland; 32, DLR, D-82234 Wessling, Germany; 33, Cornell University, Ithaca, New York 14853-1301, USA

Genome-wide survey of protein kinases required for cell cycle progression

M. Bettencourt-Dias¹, R. Giet^{1,2}, R. Sinka¹, A. Mazumdar¹, W. G. Lock¹, F. Balloux¹, P. J. Zafiroopoulos¹, S. Yamaguchi³, S. Winter³, R. W. Carthew³, M. Cooper⁴, D. Jones⁴, L. Frenz⁴ & D. M. Glover^{1,4}

¹Cancer Research UK Cell Cycle Genetics Research Group, University of Cambridge, Department of Genetics, Downing Street, Cambridge CB2 3EH, UK

²CNRS UMR 6061 Université de Rennes 1, Equipe Labellisée Ligue Nationale Contre le Cancer, Groupe Cycle Cellulaire, IFR97, 2 avenue du Pr Léon Bernard, 35043 Rennes, France

³Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Chicago, Illinois 60208, USA

⁴Polgen Division, Cyclacel Ltd, Babraham Science Park, Babraham, Cambridge CB4 2AT, UK

Cycles of protein phosphorylation are fundamental in regulating the progression of the eukaryotic cell through its division cycle. Here we test the complement of *Drosophila* protein kinases (kinome) for cell cycle functions after gene silencing by RNA-mediated interference. We observed cell cycle dysfunction upon downregulation of 80 out of 228 protein kinases, including most kinases that are known to regulate the division cycle. We find new enzymes with cell cycle functions; some of these have family members already known to phosphorylate microtubules, actin or their associated proteins. Additionally, depletion of several signalling kinases leads to specific mitotic aberrations, suggesting novel roles for familiar enzymes. The survey reveals the inter-digitation of systems that monitor cellular physiology, cell size, cellular stress and signalling processes with the basic cell cycle regulatory machinery.

The high degree of evolutionary conservation of protein kinases and their signalling pathways points to the importance of protein phosphorylation in all aspects of biology. Multiple cyclin-dependent kinases (Cdks) regulate the G1, S and G2 phases of the cell cycle in metazoans¹. The Cdks are themselves regulated and cooperate with other protein kinases; for example p34^{cdc2} (Cdk1) is inhibited by Wee1 phosphorylation, activated by Cak and acts in concert with the Polo and Aurora kinases during mitosis^{2–4}. Our limited knowledge of how cell cycle kinases are activated suggests that our understanding of these phosphorylation networks is far from complete.

Genetic studies on model organisms, particularly fungi and *Drosophila*, have proven powerful in identifying cell cycle kinases and assessing their functions. The availability of a fully sequenced genome for *D. melanogaster*, combined with the efficiency of RNA-mediated interference (RNAi) in cultured *Drosophila* cells, allows the exploration of that part of the genome not amenable to classical genetics^{5–8}. Moreover, the low redundancy of protein kinases in the *Drosophila* genome and the complete representation of their sub-families in humans make it an attractive system to screen for cell cycle functions⁹.

One-third of the kinome affects cell cycle progression

Our strategy was to transfect double-stranded RNA (dsRNA) for each of the predicted 228 kinase genes (see Supplementary Table 1) into S2 cells and monitor the effect 72 h later, a time sufficient to deplete most cell cycle proteins and reveal cellular phenotypes^{10,11} (see Methods and Supplementary Fig. 1). Results from flow cytometry screening showed delays in progression through specific cell cycle stages (which in some cases were associated with aneuploidy, polyploidy or cell death) upon downregulation of 42 protein kinases (18% of the kinome). These defects fall into four broad clusters (Fig. 1a–d) if effects on cell size are also taken into account; this parameter is typically used when defining phenotypes of cell division cycle (cdc) mutants in the yeasts.

However, flow cytometry does miss some mitotic defects. For example, RNAi on *aurora A* (*aurA*), a gene that has well defined functions in centrosomal and spindle assembly, did not reveal a phenotype by flow cytometry. This is probably because cultured

Drosophila cells are tolerant of both extra numbers of centrosomes⁶ and the complete absence of centrosomes¹². We therefore carried out RNAi on the 228 kinases and blindly quantified 20 parameters including centrosomal, spindle and chromosomal defects, the proportions of cells in the classical mitotic stages and the mitotic index (Fig. 2, see Supplementary Tables 3 and 4). Kinases were ranked according to each of their phenotypic scores (Fig. 3a–c). We defined an RNAi phenotype only when the phenotypic score was significantly different from controls in two independent experiments. According to this definition, 60 kinases showed a mitotic phenotype (Fig. 3d).

In total 80 kinases were associated with defects in cell cycle progression and/or mitosis (Figs 1 and 3). These enzymes were grouped according to their phenotype and/or functional information from other systems (Table 1 and Supplementary Table 5). Protein kinases known to have regulatory cell cycle functions (21 enzymes, indicated in Table 1) showed functions similar to corresponding *Drosophila* mutants or studies in other organisms, validating the approach.

Signal transduction, stress response and the cell cycle

Depletion of a number of protein kinases known to respond to growth factors and environmental stress, including members of NF- κ B, JNK/p38 and JAK/STAT signalling pathways, led to cell cycle defects, indicating that extracellular conditions bear directly on cell cycle progression. In response to RNAi, one cluster of these kinases showed an increased number of cells in G1 with no significant change in cell size (Table 1, group 1a). Within this cluster are PK92B and Licorne (Lic), two stress response enzymes in MAPK pathways (Table 1). In mammals, depending on the cell type, p38 MAPKs can function either to stimulate or inhibit cell proliferation through regulation of cyclin D expression¹³. Another enzyme present in this cluster is Doa, a LAMMER family kinase. Recent genetic evidence indicates that *Drosophila* Doa mutants show disrupted endoreplication of nurse cell chromosomes and fail to sustain condensation of the oocyte DNA¹⁴. Further studies will be required to determine whether this protein kinase has comparable roles in the more conventional cycles of S2 cells. Two other kinases in this group have been implicated in NF- κ B activation: JIL-1, known to regulate

chromatin structure (see Supplementary Table 5), and Pelle (Pll), the counterpart of the mammalian interleukin-1 receptor associated kinase (IRAK).

Coupling of JAK/STAT signalling to proliferation in the S2 cell line was suggested by the accumulation of cells with G1 DNA content following downregulation of the JAK kinase *hopscotch* (*hop*). Consistent with genetic interactions suggesting that *Cdk4* functions downstream of *hopscotch* (see Supplementary Table 5), we found that cells of increased size also accumulated in G1 following RNAi for either *Cdk4* (Fig. 1a) or its putative partner, *Cyclin D* (see Supplementary Fig. 2). However, *Drosophila* *Cdk4* imaginal disc clones have a longer cell cycle with no change in cell cycle profile or size distribution by flow cytometry, implicating *Cdk4* in the regulation of growth rate¹⁵. Together, this information suggests that the relative role of *Cdk4* in regulating the cell cycle may depend upon the cell type; this has also been suggested in another recent study¹⁶.

A broad spectrum of other phenotypes was observed in response to the downregulation of different signalling pathways: various mitotic phenotypes were seen for *nemo* (*nmo*) and *ik2* downregulation (Table 1, group 5), *Mkk4* downregulation resulted in chromosomal alignment defects (Table 1, group 5), and mitotic defects and/or delays in the progression through cytokinesis were observed after downregulation of several receptor-like kinases (Table 1, group 1b). It will be of interest to determine whether these phenotypes indicate other primary functions for these enzymes or secondary effects of the signalling pathways on cell cycle progression.

Nutrient sensing, cell growth and cell cycle progression

Most kinases in the cluster for which downregulation led to an increase in the proportion of small G1 cells are known members of the Tor-PDK1-S6k system (Fig. 1a and Table 1, group 2). These kinases are conserved from yeast to mammals, consistent with their known functions in sensing nutrients and regulating cell growth (see Supplementary Table 5). S6k is the effector kinase that phos-

phorylates ribosomal protein S6 to modulate translation. It can be activated by nutrient sensing through either Tor kinase or PtdIns(3,4,5)P₃-dependent kinase (PDK; Pk61C in *Drosophila*). The latter usually responds to receptor tyrosine kinase (RTK) signalling, for example insulin receptor (InR) signalling through PI(3) kinase¹⁷. The identity of the receptor tyrosine kinase in S2 cells is not clear, as *InR* RNAi itself resulted in only a weak mitotic phenotype. Downregulation of only one other kinase, *Cklα*, led to G1 delay with cells of smaller size, suggesting a new function for this enzyme in the pathway.

We also found spindle and chromosomal alignment defects following downregulation of *Gcn2*, an enzyme that phosphorylates eIF2α to impede protein translation in cells deprived of essential amino acids. Downregulation of Tor by rapamycin induces the dephosphorylation and activation of *Gcn2* (ref. 18). Thus two major pathways of nutrient control of gene expression each seem to show links not only with each other but also with cell cycle regulation, emphasizing the coordination between these processes.

Progression into and through S phase

In addition to the increase in G1 cell number in response to downregulation of known G1/S regulators (including *Cdk2* and *Cdk4*, Fig. 1a), we identified several transcriptional regulators implicated in the cell cycle and with wider functions. These include *Cdk8* and *Cdk9*, both known to phosphorylate RNA polymerase II (see Supplementary Table 5).

Delays in S phase progression indicate that CG32742 is the potential counterpart of the budding yeast protein Cdc7, a conserved kinase that phosphorylates Mcm proteins at replication origins. S phase defects coupled with lower mitotic and cytokinetic indices and increased cell death were also seen following downregulation of CG2829, the *Drosophila* counterpart of the kinase Tousled (Fig. 3d; Table 1, group 3), a conserved enzyme that regulates chromatin assembly following DNA replication and is a target of the DNA damage checkpoint (see Supplementary Table 5).

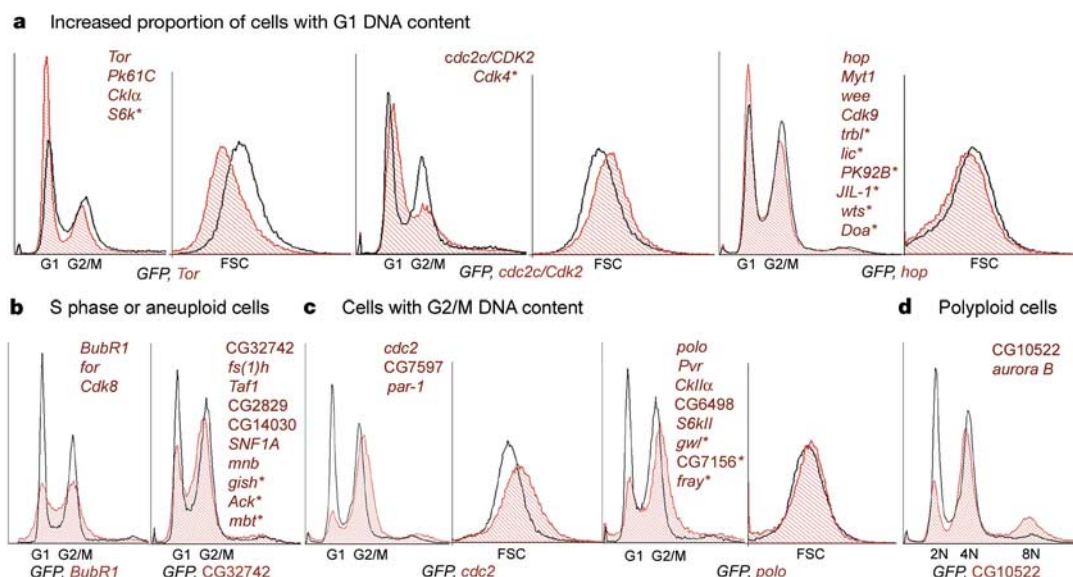


Figure 1 Cell cycle progression after RNAi of protein kinases. Each panel shows a control FACS histogram (black, cells transfected with dsRNA for green fluorescent protein (GFP)) and a FACS histogram (red) after RNAi of a kinase representative of its phenotypic class (the specific kinase is indicated under each panel, phenotypic class members listed on panel). The forward light scatter (FSC) profile reflects cell size. **a**, RNAi resulting in an increase in the proportion of cells in G1 can be associated with a reduction (left two panels), increase (middle two panels) or no significant change in cell size (right two

panels). **b**, RNAi of kinases resulting in increased proportions of cells with intermediate DNA content (S phase or aneuploid cells). These kinases have been subdivided according to the extent of accumulation of cells in the G2 phase (left versus right panel). **c**, RNAi of kinases resulting in increased proportions of cells in G2/M phase may or may not be associated with an increase in cell size (compare the two FSC histograms). **d**, RNAi resulting in an increase in the number of polyploid cells. Names followed by an asterisk indicate kinases for which the RNAi phenotype is weaker.

The observed defects are consistent with the *tousled* mutant phenotype: embryos of *tousled* show arrest of cell cycle progression in interphase, followed by apoptotic cell death¹⁹.

Protein kinases inhibiting or promoting G2/M transition

Our identification of the known major genes that regulate the G2/M transition (Table 1, group 4) provided additional validation of our screening method. Knockdown of the major mitotic kinase *cdc2/Cdk1* led to the expected increase in large G2 cells (Fig. 1c). We also identified the Cdk1-inactivating kinases *Dwee1* and *Myt1*, and the Tribbles kinase that induces proteolysis of String (*cdc25*), the Cdk1-activating protein phosphatase (see Supplementary Table 5). Downregulation of this group of kinases accelerated the G2 phase, thus shifting more cells into G1 (Fig. 1a). Downregulation of the *wt5/Lats* tumour suppressor kinase, another negative regulator of Cdk1, also led to an increase in G1 cells following RNAi. Downregulation of *S6kII* led to an increase in G2/M cells, in agreement with reports that its counterpart, the *Xenopus* kinase *p90^{rsk}*, inactivates *Myt1* during oocyte maturation²⁰. We also place a *cdc2*-related kinase, CG7597, into this category because its downregulation resulted in a low mitotic index (Table 1) and increased numbers of larger G2 cells (Fig. 1c).

New G2 functions were identified for *Taf1* and *Fs(1)h* kinases, which have previously been shown to be transcriptional regulators and are likely to be chromosomally associated because they contain bromodomains. Indeed, it has recently been reported that *Taf1* is required for transcriptional activation of the *string* (*cdc25*) gene (ref. 21). One possible human counterpart of *Fs(1)h* is *Brd4* which is suggested to be required for G2/M progression; another is *Brd2/RING3*, which participates in the transactivation of promoters dependent on E2F (see Supplementary Table 5). In genetic agree-

ment, *Drosophila* E2F1 has been shown to modulate the expression not only of genes required for G1/S but also of *string*²².

Unexpectedly, downregulation of the Pvr receptor tyrosine kinase led to an increase in the number of cells in the G2 phase (Fig. 4k, l) and the number of cells positive for Cyclin A and B (not shown), and resulted in a low mitotic index (Fig. 3d). Pvr is the counterpart of

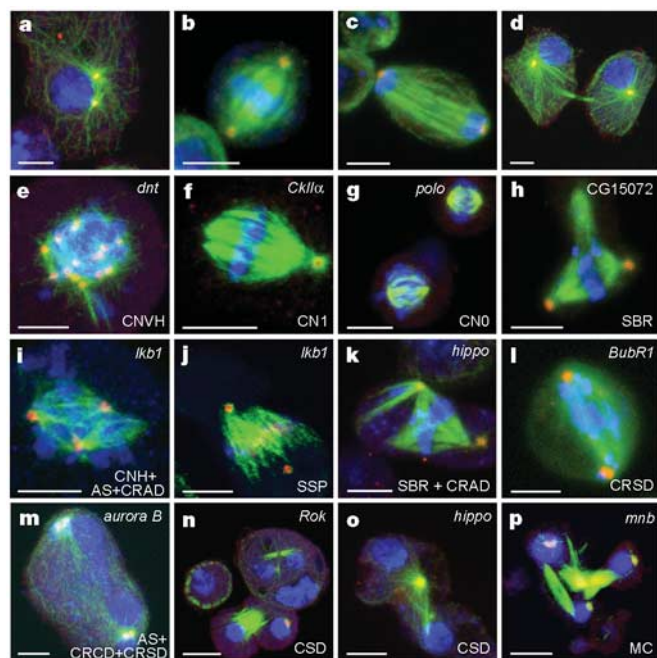


Figure 2 Examples of mitotic phenotypes seen following downregulation by RNAi of selected protein kinases. **a–d**, Control cells at **(a)** prophase, **(b)** metaphase, **(c)** late anaphase or **(d)** cytokinesis, stained to reveal α -tubulin (green), γ -tubulin (red) and DNA (blue). **e–p**, Selected RNAi phenotypes (name of gene in top right corner) illustrating some scored parameters (bottom right corner). AS, abnormal spindle; CN0, no γ -tubulin at poles; CN1, only one pole with γ -tubulin; CNVH, centrosome number very high; CRAD, chromosome alignment defects; CRCD, chromosome condensation defects; CRSD, chromosome segregation defects; CSD, central spindle defects; MC, multiple cytokinesis; SBR, branched spindle; SSP, splayed spindle. Scale bar is 5 μ m.

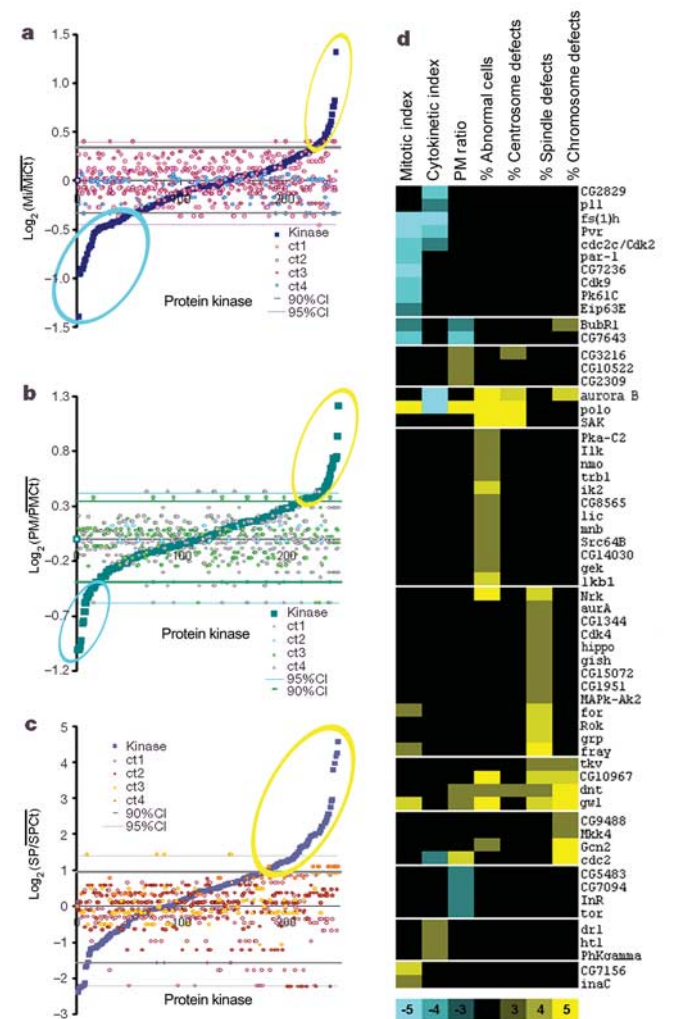


Figure 3 Quantitative analysis of mitotic RNAi phenotypes. **a–c**, Ranking of phenotypic scores (PS; filled squares) for three of the scored categories of mitotic phenotype. PS were obtained after normalization of each quantitative RNAi parameter in relation to the average of control values for each experiment (see Supplementary Methods). Filled circles represent normalized control values (ct). The scored parameters presented are **(a)** mitotic index (MI); **(b)** ratio of cells in prometaphase and metaphase versus total number of mitotic cells (PM); and **(c)** percentage of spindle abnormalities (see Supplementary Tables 3 and 4 for more details). Confidence intervals (CI) were defined on the basis of control values (see Supplementary Methods). The phenotypic score for the majority of kinases fell along a gentle slope, within the error limits for the data measurements. At the extremes, the phenotypic score was either significantly higher (encircled in yellow) or lower (encircled in blue) than controls. The mitotic parameters were scored in repeat RNAi experiments for all kinases and showed a significant correlation for each of the different variables in both experiments. **d**, Kinases with mitotic phenotypes after RNAi. Only kinases showing PS values outside the 90% CI in each of two independent experiments were deemed to have a mitotic phenotype. Each row shows the phenotype of a single kinase. Scored parameters are shown in different columns, the strength of the phenotype is shown in different colours: the extreme arbitrary values -5 and $+5$ indicate PS values outside the 99% CI at the lower or higher boundary, respectively; -4 and $+4$ are for PS values outside the 95% CI, and -3 and $+3$ are for PS values outside the 90% CI. Black boxes indicate PS values within the 90% CI.

mammalian PDGF and VEGF receptors and signals border cell migration in oogenesis, a role that it shares with EGFR (see Supplementary Table 5). RNAi against one of the Pvr ligands (*Pvf2*), but not two others (*Pvf3* and *Pvf1*), resulted in a phenotype similar to that of the receptor (Fig. 4I). This suggests that S2 cells autoregulate proliferation through a signalling pathway effective at G2, and seems at odds with the generally accepted view that

extracellular signalling directs cells through G1. However, *string* (*cdc25*) expression is highly regulated during *Drosophila* development. For example, wing disc cells spend an increasing proportion of time in G2 as they develop, and differentiating photoreceptor cell preclusters trigger increased levels of *string* expression in neighbouring cells²³. Thus G2 delay following downregulation of Pvr signalling could reflect a specific property of insect cells. It might also

Table 1 *Drosophila* kinases with cell cycle phenotypes and their putative orthologues

FlyBase name	Putative orthologues*	Previously known functions†	RNAi phenotype in current study‡
Signal transduction and stress response (Group 1a)			
<i>Pk92B</i>	HS-ASK1/MEKK5	Activates Jun in cytokine and stress-induced apoptosis (HS)	G1+
<i>lic</i>	HS-MAP2K3/6	Phosphorylates p38MAPK (HS); asymmetric development of the egg (DM)	G1+, ABN(3), SP(2)
<i>Doa</i>	HS-CLK2/3/4	Lammer dual-specificity kinase 2 (HS); meiotic progression (DM)	G1+
<i>JIL-1</i>	HS-RPS6KA5/4	Phosphorylates histone H3, activates NF-κB (HS); chromatin structure (DM)	G1+
<i>hop</i>	HS-JAK2/3/1	JAK-STAT signalling (HS); proliferation, interacts genetically with <i>CDK4</i> (DM)	G1+
Signal transduction and stress response (Group 1b)			
<i>pll</i>	HS-IRAK1	Activation of NF-κB and MAPK pathways (JNK/p38) (HS); immunity (DM)	CYT(−4)
<i>tor</i>	HS-RET1	Mutants have disrupted anterior/posterior axis, activates ras and STAT (DM)	PM(−3)
<i>drl</i>	HS-RYK	RTK (HS); axon pathfinding, Wnt receptor signalling pathway (DM)	CYT(3)
<i>htl</i>	HS-FGR2/3/1/4	FGF receptor (HS); interacts with <i>ras</i> , cell migration, upstream of <i>pbl</i> (DM)	CYT(3)
Cell growth/G1 (Group 2)			
<i>Tor</i>	HS-FRAP1	Regulates G1/S transition (HS); mutant cells are smaller and arrest in G1 (DM)	G1+, size−, CYT(−2)
<i>Pk61C</i>	HS-PDK1	Activates p70S6K (HS); upstream effector of S6k, mutant cells are smaller (DM)	G1+, size−, MI(−4)
<i>S6k</i>	HS-RPS6KB1/2	Cell proliferation and growth (HS); interacts with Pk61C and Tor, mutant cells are smaller (DM)	G1+, size−
<i>Cklα</i>	HS-CSNK1A1	Inhibits JNK cascade, Armadillo degradation, induced after DNA damage (DM)	G1+, size−
<i>InR</i>	HS-IGF1R	Signals to MAPK/ras and PI3K (HS); mutants are long-lived and have smaller body (DM)	PM(−3)
G1/S and S (Group 3)			
<i>cdc2c§</i>	HS-CDK2	G1/S and S phase progression (HS, DM)	G1+, size+, MI(−4), CYT(−3)
<i>Cdk4§</i>	HS-CDK6/4	G1/S transition (HS); cell growth (DM)	G2/M−, S+, size+, SP(3), SBR)
<i>Cdk8§</i>	HS-CDK8	Regulates RNA polymerase 2 (DM)	S+, G1 = G2/M
<i>Cdk9§</i>	HS-CDK9	Regulates RNA polymerase 2 (DM)	G1+, MI(−4)
<i>CG32742§</i>	HS-cdc7	DNA replication (HS)	S+, G2/M+
<i>CG2829§</i>	HS-TLK1/2	Chromatin assembly (HS); nuclear division, chromatin assembly, cell viability (DM)	S+, G2/M+, CYT(−4), MI(−2)
<i>SNF1A</i>	HS-AMPK2; SC-SNF1	Metabolic stress response, regulation of pol II and initiation of meiosis (SC)	S+, G2/M+, ABN (2), CN (2)
G2/M transition (Group 4)			
<i>cdc2§</i>	HS-CDK1	G2 to M phase transition, multiple mitotic functions (DM, SP, HS)	G2/M+, size+, CYT(−3), PM(4), CHR(5)
<i>Myt1§</i>	HS-Myt1	Negative regulator of CDK1 (HS); regulates mitotic entry (DM)	G1+
<i>wee1§</i>	HS-Wee1	Phosphorylates CDK1 (HS); Wee1p phosphorylates Cdc2p on Tyr 15 (SP)	G1+
<i>trb1§</i>	HS-trb2/1/SKIP3	SKIP3 upregulated in tumours (HS); induces proteolysis of String (<i>cdc25</i>) (DM)	G1+, ABN(3), SP(2)
<i>wts§</i>	HS-LATS1/2	Inhibits G2/M and promotes apoptosis (HS); interacts with <i>CycA</i> and <i>cdc2</i> (DM)	G1+
<i>S6kl</i>	HS-RSK2/1/3/6	Inactivates Myt1 (<i>Xenopus laevis</i>)	G2/M+
<i>CG7236</i>	HS-CDKL1	Involved in gliosis (HS)	MI(−5), multinucleate cells
<i>CG7597</i>	HS-CRK7; CE-B0285.1	MPM-2 antigen (HS); RNAi <i>in vivo</i> results in slow growth (CE)	G2/M+, size+, MI(−2)
<i>Eip63E</i>	HS-PFTAIR-1	Embryonic and larval development (DM)	MI(−3)
<i>Taf1</i>	HS-Taf11250	Associated with TATA box binding protein, induces G1 progression through p53 (HS); transcriptional activation of <i>string/cdc25</i> (DM)	S+, G2/M+, MI(−2), ABN(2)
<i>fs(1)h</i>	HS-BRD4/MCAP or RING3	BRD4 associates with chromosomes, G2/M function, RING3 trans-activates genes dependent on E2F (HS)	S+ and/or aneuploidy, G2/M+, MI(−5), CYT(−5), PM(2)
<i>Pvr</i>	HS-VEGFR1/2/3	Proliferation and cell migration (HS); organizes actin cytoskeleton (DM)	S−, G2/M+, MI(−5), CYT(−4)
<i>par-1</i>	HS-MARK3	Phosphorylates CDC25C (HS); interacts genetically with <i>lkb1</i> to regulate polarity (DM)	G2/M+, size+, MI(−4), ABN (2)
<i>Ack</i>	HS-Ack1	Effector of <i>cdc42</i> (HS); dorsal closure, expressed in mitotic domains (DM)	S+, G2/M+
Mitosis (Group 5)			
<i>polo§</i>	HS-plk1; SC-cdc5	Multiple mitotic functions (HS)	G2/M+, MI(5), CYT(−5), PM(5), ABN(5), SP(2), CN(5:CN0)
<i>SAK§</i>	HS-SAK	Required for mitosis (<i>Mus musculus</i>)	ABN(5), CN(5), AS
<i>aurA§</i>	HS-Aurora A	Entry into mitosis (HS); defects in centrosome maturation and spindle formation (DM)	SP(3)
<i>Cklα</i>	HS-CK2A1	Phosphorylates p53, multiple signalling pathways (HS); circadian clock (DM)	G2/M+, ABN(2: CN1, CRLC)
<i>CG7156</i>	HS-RSK-LIKE	Novel RSKL similar to JIL, S6K and S6KII (HS)	G2/M+, MI(4), SP(2)
<i>Pka-C2</i>	HM-PKA-Cbeta1; SC-PKA1 or 2	Regulates mitotic progression through <i>cdc20</i> (SC)	ABN(3), SP(2), CHR(2)
<i>lkb1</i>	HS-LKB1	Tumour suppressor, activates 13 kinases of the AMPK subfamily (HS); oocyte microtubule organization (DM)	ABN(4), CN, SP

exemplify wider possibilities for the regulation of G2 progression by external signalling. Indeed, recent characterization of the mouse MKK7 knockout phenotype also suggests that signalling through the JNK pathway couples environmental cues to G2/M regulation²⁴.

Lkb1 signalling has pleiotropic roles in the cell cycle

Our screen has identified new roles for several members of the Lkb1

protein kinase cascade. Overexpression of wild-type (but not kinase-inactive) LKB1 can suppress the growth of some human cancer cell lines, apparently through p53-mediated expression of the p21 Cdk inhibitor²⁵. It has recently been shown that LKB1 can activate some 13 members of the AMPK subfamily²⁶. We found cell cycle phenotypes upon downregulation of *lkb1* or three putative Lkb1 targets, CG15072, *SNF1A* and *par1*. Downregulation of either

Table 1 – (contd)

FlyBase name	Putative orthologues*	Previously known functions†	RNAi phenotype in current study‡
Mitosis (contd)			
CG15072	HS-KIAA0999/QSK	AMPK-related kinase activated by LKB1 (HS)	SP(3), CRAD
<i>nmo</i>	HS-NemoLK	Wnt signalling, polarization/rotation of cells, interacts with <i>NF-κB</i> (DM)	ABN(3:SP(2), CN, CRLC)
<i>ik2</i>	HS-TBK1	NF-κB signalling (HS, DM); defence response (DM)	ABN(4), SP(2)
<i>inaC</i>	SC-PKC1	Mutants have behaviour defects (DM); morphogenesis checkpoint (SC)	MI(3)
<i>dnt</i>	HS-RYK	Upregulated in ovarian cancer (HS); interacts with <i>drl</i> (DM)	PM(3), ABN(3), CN(3), SP(3), CHR(5)
<i>for</i>	HS-PRKG1	NO/cGMP/cGK signalling, negative regulator of cell proliferation (HS); response to hypoxia, feeding behaviour (DM)	S+, aneuploidy, MI(3), ABN(2), SP(4: SMO)
CG3216	HS-Atrial natriuretic peptide receptor1‡	Responds to cGMP; inhibits proliferation (HS)	PM(3), CN(3)
CG1951	HS-KIAA1360/NTKL‡; SC-SCY1	NTKL localizes to centrosomes during mitosis (HS)	SP(3)
CG6498	HS-MAST1 or 2	Localizes to spermatid manchette, activates NF-κB (<i>Mus musculus</i>)	G2/M+, CRAD
<i>Mkk4</i>	HS-MAP2K4	JAK/STAT and JNK cascades, links stress response to cell cycle (HS)	CHR(3: CRAD)
<i>MAPK-Ak2</i>	HS-MAPKAPK2	Activated in response to IFN in the p38 pathway (HS)	SP(3), CRAD
<i>fray</i>	HS-OSR1 or SPAK	Oxidative stress response, phosphorylates PAK1 (HS); nerve ensheathment (DM)	MI(3), SP(5), CRAD
<i>mbt</i>	HS-PAK7/5/4	Interacts with Cdc42 and Rac (HS); cytoskeleton and photoreceptor development (DM)	S+ and/or aneuploidy, G2/M+, CRAD
<i>llk</i>	HS-ILK1/2; CE-ILK	Linkage of integrins to actin cytoskeleton (HS); focal adhesions of cytoskeleton (DM)	ABN(3: CHR(2: CRLC), AS
<i>Src64B</i>	HS-FYN‡	Regulates actin polymerization, cell proliferation (DM)	MI(–2), ABN (3: SP, CHR)
<i>gek</i>	HS-CDC42BP	Abnormal accumulation of F-actin in oogenesis (DM)	ABN(3: AS, CRAD)
CG1344	HS-Pace-1	Cell spreading and motility, colocalizes with ezrin in lamellipodia (HS)	SP(3)
<i>gish</i>	HS-CK1G3; CE-Y106G6E.6; SC-YCK1 or 2	Glial cell migration (DM); mitotic spindle orientation (CE); growth and division, cell morphogenesis and cytokinesis (SC)	S+ and/or aneuploidy, G2/M+, MI(2), SP(3)
<i>gwl</i>	HS-FLJ14813; SC-rim15; SP-cek1‡	Sporulation and meiosis (SC); cek1 suppresses cut8 (SP); chromosome condensation defects (DM)	G2/M+, MI(4), PM(3), ABN(5), SP(4), CHR(5, CRSD)
<i>mnb</i>	HS-Dyrk1; CE-mbk1/2‡	Candidate target of Down's syndrome (HS); mutants have small brains (DM); spindle positioning and asymmetric cell division (CE)	S+ and/or aneuploidy, G2/M+, ABN(3: AS)
CG2309	HS-ERK8	Activated by SRC (HS)	PM(3)
CG10967	HS-ULK2; CE-Unc-51	Axon morphogenesis and elongation, might signal through ras (CE)	ABN(5), SP(4), CN(2), CHR(4)
<i>Gcn2</i>	HS-KIAA1338/GCN2	Phosphorylates eIF2α during amino acid deprivation (HS); protein synthesis in stress response (SC)	ABN(3), SP(2), CHR(5: CRAD)
<i>tkv</i>	HS-BMPR1B	Type I TGF-β receptor (HS); cell growth and division, anterior/posterior patterning (DM)	ABN(2), SP(3), CHR(3: CRAD, polyploid cells)
<i>Nrk</i>	HS-MUSK	Muscle-specific tyrosine kinase receptor (HS); interacts with <i>ras</i> in oocytes (DM)	ABN(5), SP(4), CHR(2, CRAD)
CG8565	HS-SRPK2‡; CE-SPK-1‡	Pre-mRNA splicing (HS); SPK-1 is required for embryogenesis and germline development (CE)	ABN(3), CN (2: CN1)
CG9488	HS-DDR2‡	Extracellular matrix remodelling (HS)	CHR(3)
Checkpoints (Group 6)			
<i>BubR1</i> §	HS-BubR1; SC-Bub1	Spindle assembly checkpoint (HS); mutants have low mitotic index and premature mitotic exit (DM)	Aneuploidy, MI(–3), PM(–3), ABN(2: CSD), CHR(3: CRLC)
CG7643§	HS-TTK; SC-MPS1	Spindle assembly checkpoint and centrosome duplication (HS); duplication of SPB and spindle assembly checkpoint (SC)	MI(–4), PM(–4), SP(2)
CG14030/ <i>Bub1</i> §	HS-BUB1‡; SC-Bub1	Spindle assembly checkpoint (HS); does not contain KEN box, functionally similar to HS-Bub1 and SC-Bub1 (DM)	Aneuploidy, ABN(3: CRLC, CSD)
<i>Grp</i> §	HS-Chk1	Replication and DNA damage (G2) checkpoint (HS); cell cycle coordination in syncytial embryo, mutant has defects in mitotic entry (DM)	SP(4), CRAD, CRLC
Telophase and cytokinesis (Group 7)			
CG7094	HS-CSNK1A1‡	Wnt signalling (HS)	PM(–3)
CG5483	HS-KIAA1790	Similar to leucine-rich repeat kinase (LRRK1) (HS)	PM(–3), CN(2)
<i>Hippo</i>	HS-STK4/3; SC-cdc15‡	Apoptosis (HS); apoptosis and cell cycle exit (DM); mitotic exit network (SC)	SP(3), CRLC
<i>aurora B</i> §	HS-AurB‡; SC-lp1p‡	Spindle assembly checkpoint and cytokinesis (HS); chromosome condensation and cytokinesis (DM)	8N peak, CYT(–5), CN(4), ABN(5), CHR(4)
CG10522§	HS-CIT	Cytokinesis (HS, DM)	8N peak, PM(3), ABN(2)
<i>Rok</i> §	HS-ROCK	Cytokinesis (HS, SP); tissue polarity (DM)	ABN(2), SP(4: CSD)
<i>Phkγ</i>	HS-PHKG1	Metabolism (HS); embryonic morphogenesis (DM)	CYT(3), CSD, CRLC

The table shows 80 protein kinases, grouped on the basis of RNAi phenotypes (this study) and/or functional information from other systems.
*We obtained orthologues from the Inparanoid database⁴⁹ (confidence value = 0.05 or higher). Putative human (HS) homologues and potential counterparts from *C. elegans* (CE), *S. cerevisiae* (SC) and *S. pombe* (SP) are suggested for cases where known phenotypes are helpful in assessing function.
†Known functions are described for the *Drosophila* (DM) kinases, or for human, *C. elegans*, *S. pombe* or *S. cerevisiae* orthologues. Additional information, references and sources of information relating to the functions of orthologues for individual proteins are listed in Supplementary Table 5.
‡Descriptions of observed phenotypes. Increases (+) or decreases (–) in cell size or the proportion of cells in a given phase of the cell cycle (G1, S or G2/M) are indicated. The level of confidence for each phenotype corresponds to the scale described in Fig. 3. Phenotypic scores (PS) falling outside the confidence interval (CI) are shown as follows: ±2 (85% CI), ±3 (90% CI), ±4 (95% CI), ±5 (99% CI). ABN, all mitotic abnormalities; AS, abnormal spindle; CHR, chromosome abnormalities; CN, centrosome abnormalities; CN0, no γ-tubulin at the poles; CRAD, chromosome alignment defect; CRLC, lagging chromatids; CRSD, chromosome segregation defect; CSD, central spindle defect; CYT, cytokinetic index; MI, mitotic index; PM, ratio of cells in prometaphase and metaphase versus the total number of mitotic cells; SBR, branched spindle; SMO, monopolar spindle; SP, spindle abnormalities. Additional information can be found in Supplementary Table 4.
§Reflects a previously known cell cycle regulatory kinase.
‡‡If there is no clear orthologue from the Inparanoid database, the closest homologue from a BLAST⁵⁰ search (NCBI) is indicated.

CG15072 or *lkb1* had strong effects on spindle morphology (Fig. 2). In contrast, RNAi of the AMP-activated protein kinase *SNF1A* (Table 1, group 3) led to pleiotropic defects including an increase in number of S phase cells. This increase was also seen following downregulation of the SNF1A regulator *SNF4 γ* (see Supplementary Fig. 2). This suggests a direct link between the sensing of cellular energy levels (by AMP-regulated protein kinase) and cell cycle progression. Downregulation of *par1* resulted in a striking increase in the number of G2 cells. Because Par1 is better known as an enzyme that cooperates with Lkb1 to regulate cellular polarity, this highlights the need for further studies of this network in cell cycle progression.

Mitotic functions

Among the enzymes for which depletion led to mitotic defects was the well characterized Polo kinase. RNAi of *polo* led to the typical features of strongly hypomorphic *polo* mutants²⁷: a marked increase in metaphase-arrested cells (Fig. 3d) and a tenfold increase in spindles with no γ -tubulin at the poles (Figs 2 and 3d). These defects reflect the role of Polo in regulating centrosome maturation and the metaphase–anaphase transition^{4,27}. The Aurora A kinase also falls into this group, as do several other kinases showing equal or greater RNAi spindle defects. Many of these kinases have not previously been studied in *Drosophila*, and our attempts to find mammalian counterparts by sequence homology also identified poorly characterized kinases (Table 1). Of these, the CG1951 and CG6498 kinases are particularly interesting, because their putative mammalian counterparts are associated with centrosomes and with the manchette microtubules of spermatids (Table 1, group 5; see Supplementary Table 5). RNAi on *CkII α* led to increased numbers of G2/M cells and mitotic defects, including spindles with a single

centrosome (Fig. 2). An increase in centrosomal abnormalities was also observed with RNAi of the CkII α regulator *CkII β* (not shown). Although this may indicate a direct mitotic function, the known pleiotropy of CkII (ref. 28) makes it difficult to exclude indirect effects.

We also found mitotic defects as a result of downregulating two Ste20-related kinases: abnormal spindles and abnormal chromosome behaviour were observed with RNAi of *fray* (Fig. 4h, i), and an increase in G2/M cells, and possibly aneuploidy, were observed following knockdown of *mushroom bodies tiny* (*mbt*). The Mbt kinase has been shown to localize to adherens junctions in a cdc42-GTP dependent manner²⁹. It is not clear what the precise vertebrate counterpart of Mbt is, but one possible orthologue, PAK5, regulates both the actin and tubulin cytoskeletons³⁰.

The role of the actin cytoskeleton in microtubule attachment to kinetochores³¹ and early mitotic events (such as spindle positioning and assembly³²) has only recently become apparent. We found evidence for a role of the actin cytoskeleton in mitosis based on RNAi of the putative actin cytoskeleton regulators *integrin-linked kinase* (*Ilk*), *Src64B* and *genghis kahn* (*gek*) (Table 1, group 5). Knockdown of *gek*, an effector of cdc42 known to regulate actin polymerization in the developing egg chamber³³, led to the formation of abnormal spindles with chromosome alignment defects (Table 1, group 5).

Finally, defects in spindle morphology and chromosome congression and/or segregation following RNAi of *greatwall* (*gwl*) suggested new mitotic functions for this kinase (Figs 3d and 4b, c). Spindles of metaphase length contained uncongressed chromosomes, whereas cells with elongated anaphase-like spindles had unequal numbers of chromosomes segregated to the poles after *gwl* RNAi (Fig. 4c). To determine whether lagging or pole-associated

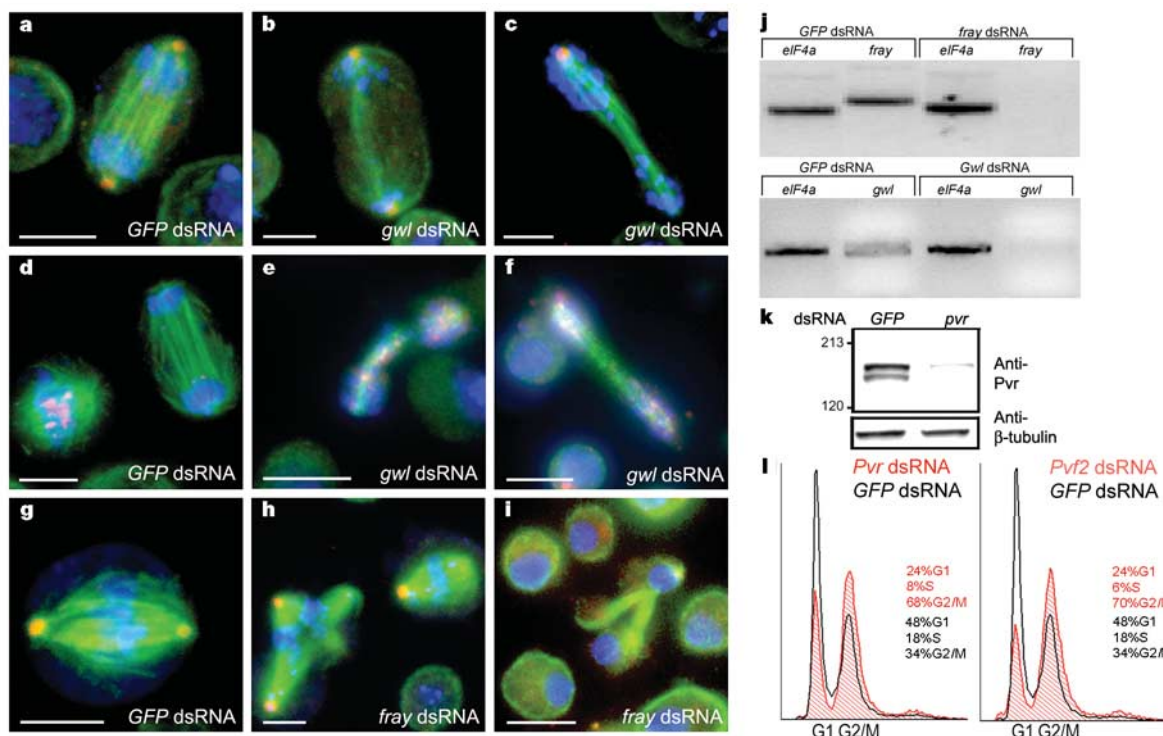


Figure 4 New cell cycle roles for Gwl, Fray and Pvr kinases. α -tubulin (green), γ -tubulin (red) and DNA (blue) staining of control cells treated with dsRNA for GFP (a, g). b, c, RNAi of *gwl* leads to chromosome segregation and spindle abnormalities. Note the unequal amounts of chromatin at the spindle poles. d, In control cells, MEI-S332 (red) is lost from centromeres after metaphase. e, f, After *gwl* RNAi, MEI-S332 staining is associated with chromosomes towards the centre of the spindle (e) or at the poles (f) of anaphase-like

spindles. h, i, RNAi of *fray* leads to severe spindle defects. j, RNAi of *fray* and *gwl* leads to reduction of RNA monitored by RT-PCR. RNA levels for *elF4a* are shown as a control. k, RNAi of *Pvr* leads to reduction of Pvr protein. l, *Pvr* RNAi leads to an increase in cells with G2 DNA content (shown in red; control in black) and the *Pvr* ligand, encoded by *Pvf2*, shows the same phenotype. Scale bar is 5 μ m.

chromosomes were separated chromatids, we examined the distribution of the Mei-S332 protein³⁴. In control cells, Mei-S332 is lost from centromeres as sisters separate at the metaphase–anaphase transition (Fig. 4d). In contrast, Mei-S332 is not lost from centromeres in comparable *gwl* RNAi cells (Fig. 4e, f). The function of Gwl might therefore be to regulate the attachment of sister kinetochores to opposite spindle poles to enable sister separation, to break sister chromatid cohesion, or perhaps both functions. However, we did not observe the pronounced chromosome condensation defects that have been described in *Drosophila greatwall* mutants³⁵.

Spindle integrity checkpoint

The spindle integrity checkpoint delays anaphase until all chromosomes are correctly aligned with sister kinetochores attached to opposite poles and under tension³⁶. Its failure leads to premature anaphase, and therefore to a lowered mitotic index with lagging chromatids^{36,37}. Our survey identified such phenotypes after RNAi of the spindle integrity checkpoint kinases *BubR1* (ref. 38) and CG7643, the *Drosophila* counterpart of Mps1 kinase (Figs 2, 3d and Table 1, group 6). Surprisingly, depletion of the *Bub1* checkpoint kinase³⁸ led to no change in mitotic index or of the proportion of cells passing through metaphase. *Bub1* RNAi also does not compromise anaphase timing in mammalian cells³⁹; this is consistent with the observation that BubR1 and Mps1, but not Bub1, dynamically exchange from the kinetochore to delay anaphase onset⁴⁰.

The report that Mps1 is also required for centrosome replication⁴¹ in human cells is controversial⁴². We saw no indication of this requirement following CG7643 RNAi in S2 cells, but as noted above, these cells tolerate considerable variation in centrosome number^{6,12}. If Mps1 is required for centrosome duplication in some aspect of *Drosophila* development, the requirement is not seen in this cell line.

Late mitosis and cytokinesis

Within this group of kinases, Hippo, a recently characterized regulator of apoptosis and cell cycle exit⁴³, showed notable spindle and central spindle defects upon downregulation (Fig. 2). We also identified the major kinases already known to regulate cytokinesis (Table 1, group 7). These include the passenger kinase Aurora B¹¹ as well as two enzymes that phosphorylate the myosin regulatory light chain: the Rho-dependent and Citron kinases^{44,45}. Downregulation of *Rho*-kinase led to central spindle defects (Fig. 2) with no increase in the number of polyploid cells, suggesting that cells recover and complete cytokinesis. Depletion of the Citron kinase (CG10522) resulted in the formation of many binucleate cells (Fig. 1d), as previously reported⁴⁵.

Conclusions

Our study has identified new cell cycle protein kinases and assigned new cell cycle functions to previously known enzymes. The G2 delay seen following downregulation of the PDGF/VEGF-related receptor, Pvr, exemplifies one such new role. The survey further highlights those aspects of cellular physiology regulated by protein phosphorylation that are intimately linked to cell cycle progression. These include external signalling from growth factors or nutrients, cellular responses to stress and regulation of cell growth. We have also found new mitotic functions for enzymes predicted to regulate cytoskeletal elements; those that link extracellular signalling and actin cytoskeleton regulation with the G2/M transition and mitosis are of particular interest. Further studies of these kinases should shed more light on these and similar findings by others^{24,31,32}. Furthermore, the assays developed here (see Supplementary Table 4) and the phenotypes identified might be used as a platform for identifying interacting genes.

Although we adopted conservative criteria, we identified most previously known cell cycle kinases. We found phenotypes consistent with equivalent mutants in the fruitfly and other organisms. This validates our approach and gives us confidence that the

approach has identified the great majority of kinases that regulate cell cycle progression in S2 cells. However, the ability of this cell line to tolerate defects such as abnormal centrosome numbers means that we may have overlooked kinases that are absolutely essential in the whole organism. We were, for example, unable to assign a cell cycle function to the *Drosophila* counterpart of the human Nek2 kinase. Only when we carefully examined this RNAi phenotype in separate experiments were we able to detect a very weak phenotype affecting centrosome integrity⁴⁶. Nevertheless, the low degree of redundancy in the *Drosophila* genome does facilitate identification of most cell cycle functions, and the high level of conservation of cell cycle regulators suggests that the study of human counterparts will benefit the understanding and treatment of proliferative disease. □

Methods

Double-stranded RNA synthesis

Genomic DNA or cDNA was used to make dsRNA as described previously⁴⁷; the average length of dsDNA was 500 base pairs (bp). The set of protein kinases was defined based on previous reports^{9,48} and annotation in Flybase, using homologies with protein kinase catalytic sites⁹. A list of primer pairs can be found in Supplementary Table 1. dsRNA was analysed by electrophoresis in 1.5% agarose gels for the purposes of quantification and to ensure that the RNA migrated as a single band.

Cell culture and transfections

Drosophila S2 cells (cell line derived from embryos)⁴⁷ were cultured and transfected with 10 µg of dsRNA and 10 µl of Transfast (Promega) in six-well plates as described in ref. 47 and in Supplementary Fig. 1. Cells were collected after three days.

Western blotting and RT-PCR

For protein analysis, an aliquot of the cells was resuspended and boiled in Laemmli buffer. Standard procedures were used for western blotting (see Supplementary Methods for details of antibodies used). For PCR of cDNA (RT-PCR) analysis, RNA was extracted from cells using the Qiagen Rneasy Protect Mini Kit and RT-PCR was performed using the SuperScript First Strand Synthesis System (Invitrogen) according to the manufacturer's instructions.

Immunofluorescence analysis

S2 cells were harvested 3 days after transfection, plated on glass coverslips and fixed 1 h later in 4% formaldehyde in PHEM buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 4 mM MgCl₂). Cells were permeabilized and washed using PBSTB (PBS containing 0.1% Triton X-100 and 1% BSA). DNA was stained using TOTO3-iodide (Molecular Probes) or DAPI. Vectashield mounting medium H-1200 was purchased from Vector Laboratories. Coded numbers were assigned to control and sample slides so that blind counts could be performed. A total of 1,000–3,000 cells (of which at least 60 were mitotic cells) were scored per slide, using a Zeiss Axiovert 200M microscope. Cells were categorized according to phase of mitosis and to centrosome, spindle and DNA morphology. On the basis of this categorization, defective mitotic cells were assigned to one or more of 20 categories of mitotic phenotypic abnormality (see Supplementary Table 3), coded to facilitate computer analysis of the data (see Supplementary Table 4 for downloadable datasheet). Two sets of data were obtained for each kinase, from two independent experiments. Seven phenotypic parameters (mitotic index, cytokinetic index, PM ratio (ratio of the number of cells in prometaphase and metaphase to the total number of mitotic and cytokinetic cells), percentage of mitotic defects, percentage of centrosome defects, percentage of spindle defects, and percentage of chromosome defects) were compared across the whole data set. Details of the statistical analysis can be found in Supplementary Methods. Images were acquired using a confocal scanning head (model 1024; Bio-Rad Laboratories) mounted on an Optiphot microscope (Nikon) and prepared for publication using Adobe Photoshop.

Flow cytometry

For FACS analysis, 2 ml cells were recovered 3 days after transfection and fixed in 70% ice-cold ethanol. For analysis of levels of cyclin A, B and phospho-histone H3, cells were permeabilized and blocked using PBS with 1% BSA and 0.25% Triton X-100. All incubations with antibodies and wash steps were performed using PBS with 1% BSA. The cells were incubated at 37 °C for 30 min in PBS containing 100 µg ml⁻¹ RNase (previously boiled for 5 min) and 100 µg ml⁻¹ propidium iodide before analysis. For analysis of DNA content we used a Becton Dickinson FACScan and a Becton Dickinson LSR and acquired data from 30,000 cells for each sample. Results were analysed using Summit from Dako Cytometry and Multicycle. At least three independent experiments were performed for each protein.

Received 2 July; accepted 28 October 2004; doi:10.1038/nature03160.

1. Pines, J. Cyclins and cyclin-dependent kinases: theme and variations. *Adv. Cancer Res.* **66**, 181–212 (1995).
2. Nigg, E. A. Mitotic kinases as regulators of cell division and its checkpoints. *Nature Rev. Mol. Cell Biol.* **2**, 21–32 (2001).
3. Donaldson, M. M., Tavares, A. A., Hagan, I. M., Nigg, E. A. & Glover, D. M. The mitotic roles of Polo-

- like kinase. *J. Cell Sci.* **114**, 2357–2358 (2001).
4. Barr, F. A., Sillje, H. H. & Nigg, E. A. Polo-like kinases and the orchestration of cell division. *Nature Rev. Mol. Cell Biol.* **5**, 429–440 (2004).
5. Clemens, J. C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA* **97**, 6499–6503 (2000).
6. Goshima, G. & Vale, R. D. The roles of microtubule-based motor proteins in mitosis: comprehensive RNAi analysis in the *Drosophila* S2 cell line. *J. Cell Biol.* **162**, 1003–1016 (2003).
7. Kiger, A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27 (2003).
8. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
9. Manning, G., Plowman, G. D., Hunter, T. & Sudarsanam, S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520 (2002).
10. Somma, M. P., Fasulo, B., Cenci, G., Cundari, E. & Gatti, M. Molecular dissection of cytokinesis by RNA interference in *Drosophila* cultured cells. *Mol. Biol. Cell* **13**, 2448–2460 (2002).
11. Giet, R. & Glover, D. M. *Drosophila* aurora B kinase is required for histone H3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis. *J. Cell Biol.* **152**, 669–682 (2001).
12. Debec, A. & Abbadie, C. The acentriolar state of the *Drosophila* cell lines 1182. *Biol. Cell* **67**, 307–311 (1989).
13. Nebreda, A. R. & Porras, A. p38 MAP kinases: beyond the stress response. *Trends Biochem. Sci.* **25**, 257–260 (2000).
14. Morris, J. Z., Navarro, C. & Lehmann, R. Identification and analysis of mutations in bob, Doa and eight new genes required for oocyte specification and development in *Drosophila melanogaster*. *Genetics* **164**, 1435–1446 (2003).
15. Meyer, C. A. *et al.* *Drosophila* Cdk4 is required for normal growth and is dispensable for cell cycle progression. *EMBO J.* **19**, 4533–4542 (2000).
16. Malumbres, M. *et al.* Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell* **118**, 493–504 (2004).
17. Kozma, S. C. & Thomas, G. Regulation of cell size in growth, development and human disease: PI3K, PKB and S6K. *Bioessays* **24**, 65–71 (2002).
18. Cherkasova, V. A. & Hinnebusch, A. G. Translational control by TOR and TAP42 through dephosphorylation of eIF2 α kinase GCN2. *Genes Dev.* **17**, 859–872 (2003).
19. Carrera, P. *et al.* Tousled-like kinase functions with the chromatin assembly pathway regulating nuclear divisions. *Genes Dev.* **17**, 2578–2590 (2003).
20. Tunquist, B. J. & Maller, J. L. Under arrest: cytoskeletal factor (CSF)-mediated metaphase arrest in vertebrate eggs. *Genes Dev.* **17**, 683–710 (2003).
21. Maile, T., Kwoczynski, S., Katzenberger, R. J., Wassarman, D. A. & Sauer, F. TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. *Science* **304**, 1010–1014 (2004).
22. Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. & Edgar, B. A. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93**, 1183–1193 (1998).
23. Lee, L. A. & Orr-Weaver, T. L. Regulation of cell cycles in *Drosophila* development: intrinsic and extrinsic cues. *Annu. Rev. Genet.* **37**, 545–578 (2003).
24. Wada, T. *et al.* MKK7 couples stress signalling to G2/M cell-cycle progression and cellular senescence. *Nature Cell Biol.* **6**, 215–226 (2004).
25. Tiainen, M., Vaahtomeri, K., Ylikorkala, A. & Makela, T. P. Growth arrest by the LKB1 tumor suppressor: induction of p21(WAF1/CIP1). *Hum. Mol. Genet.* **11**, 1497–1504 (2002).
26. Lizcano, J. M. *et al.* LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* **23**, 833–843 (2004).
27. Donaldson, M. M., Tavares, A. A., Ohkura, H., Deak, P. & Glover, D. M. Metaphase arrest with centromere separation in polo mutants of *Drosophila*. *J. Cell Biol.* **153**, 663–676 (2001).
28. Litchfield, D. W. Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem. J.* **369**, 1–15 (2003).
29. Schneeberger, D. & Raabe, T. Mbt, a *Drosophila* PAK protein, combines with Cdc42 to regulate photoreceptor cell morphogenesis. *Development* **130**, 427–437 (2003).
30. Cau, J., Faure, S., Comps, M., Delsert, C. & Morin, N. A novel p21-activated kinase binds the actin and microtubule networks and induces microtubule stabilization. *J. Cell Biol.* **155**, 1029–1042 (2001).
31. Yasuda, S. *et al.* Cdc42 and mDia3 regulate microtubule attachment to kinetochores. *Nature* **428**, 767–771 (2004).
32. Rosenblatt, J., Cramer, L. P., Baum, B. & McGee, K. M. Myosin II-dependent cortical movement is required for centrosome separation and positioning during mitotic spindle assembly. *Cell* **117**, 361–372 (2004).
33. Luo, L. *et al.* Genghis Khan (Gek) as a putative effector for *Drosophila* Cdc42 and regulator of actin polymerization. *Proc. Natl Acad. Sci. USA* **94**, 12963–12968 (1997).
34. Tang, T. T., Bickel, S. E., Young, L. M. & Orr-Weaver, T. L. Maintenance of sister-chromatid cohesion at the centromere by the *Drosophila* MEL-5332 protein. *Genes Dev.* **12**, 3843–3856 (1998).
35. Yu, J. *et al.* Greatwall kinase: a nuclear protein required for proper chromosome condensation and mitotic progression in *Drosophila*. *J. Cell Biol.* **164**, 487–492 (2004).
36. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421 (2003).
37. Jones, J. T., Myers, J. W., Ferrell, J. E. & Meyer, T. Probing the precision of the mitotic clock with a live-cell fluorescent biosensor. *Nature Biotechnol.* **22**, 306–312 (2004).
38. Logarinho, E. *et al.* Different spindle checkpoint proteins monitor microtubule attachment and tension at kinetochores in *Drosophila* cells. *J. Cell Sci.* **117**, 1757–1771 (2004).
39. Johnson, V. L., Scott, M. L., Holt, S. V., Hussein, D. & Taylor, S. S. Bub1 is required for kinetochore localization of BubR1, Cenp-E, Cenp-F and Mad2, and chromosome congression. *J. Cell Sci.* **117**, 1577–1589 (2004).
40. Howell, B. J. *et al.* Spindle checkpoint protein dynamics at kinetochores in living cells. *Curr. Biol.* **14**, 953–964 (2004).
41. Fisk, H. A., Mattison, C. P. & Winey, M. Human Mps1 protein kinase is required for centrosome duplication and normal mitotic progression. *Proc. Natl Acad. Sci. USA* **100**, 14875–14880 (2003).
42. Stucke, V. M., Sillje, H. H., Arnaud, L. & Nigg, E. A. Human Mps1 kinase is required for the spindle assembly checkpoint but not for centrosome duplication. *EMBO J.* **21**, 1723–1732 (2002).
43. Harvey, K. F., Pfeleger, C. M. & Hariharan, I. K. The *Drosophila* Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell* **114**, 457–467 (2003).
44. Matsumura, F., Totsukawa, G., Yamakita, Y. & Yamashiro, S. Role of myosin light chain phosphorylation in the regulation of cytokinesis. *Cell Struct. Funct.* **26**, 639–644 (2001).
45. D'Avino, P., Savoian, M. & Glover, D. Mutations in sticky lead to defective organization of the contractile ring during cytokinesis and are enhanced by Rho and suppressed by Rac. *J. Cell Biol.* **5**, 61–71 (2004).
46. Prigent, P., Glover, D. & Giet, R. The *Drosophila* Nek2 protein kinase is required for centrosome integrity and is able to regulate membrane addition during cytokinesis. *Exp. Cell Res.* (in the press).
47. Bettencourt-Dias, M., Sinka, R., Frenz, L. & Glover, D. M. in *Gene Silencing by RNA Interference: Technology and Application* (ed. Sohail, M.) (CRC Press, Washington, 2004).
48. Morrison, D. K., Murakami, M. S. & Cleghon, V. Protein kinases and phosphatases in the *Drosophila* genome. *J. Cell Biol.* **150**, F57–F62 (2000).
49. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
50. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Cancer Research UK and BBSRC for supporting work in the CR-UK Cell Cycle Genetics Research Group. The BBSRC support was part of a LINK Programme in which the Department of Trade and Industry also provide support for Cyclacel Ltd. An NIH grant supported work in R.W.C.'s group and JSPS funded S.Y. We thank N. B. Carmo, C. Malone and N. Miller for help with experiments. We are grateful to J. C. Pezzullo for his expertise with excel statistical macros and to F. Scaerou, C. Midgley and J. Pereira-Leal for discussions. We also thank T. Hunt, J. Maller and our colleagues, particularly A. Carpenter, M. Segal and M. Savoian for their comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B.-D. (mbcd2@cam.ac.uk) or to D.M.G. (dmg25@mole.bio.cam.ac.uk).

The sequence and analysis of duplication-rich human chromosome 16

Joel Martin¹, Cliff Han², Laurie A. Gordon¹, Astrid Terry¹, Shyam Prabhakar⁴, Xinwei She⁵, Gary Xie^{1,2}, Uffe Hellsten¹, Yee Man Chan⁶, Michael Altherr^{1,2}, Olivier Couronne⁴, Andrea Aerts¹, Eva Bajorek⁶, Stacey Black⁶, Heather Blumer², Elbert Branscomb^{1,3}, Nancy C. Brown², William J. Bruno², Judith M. Buckingham², David F. Callen², Connie S. Campbell², Mary L. Campbell², Evelyn W. Campbell², Chenier Caoile⁶, Jean F. Challacombe², Leslie A. Chasteen², Olga Chertkov², Han C. Chi², Mari Christensen³, Lynn M. Clark², Judith D. Cohn², Mirian Denys⁶, John C. Dettler¹, Mark Dickson⁶, Mira Dimitrijevic-Bussod², Julio Escobar⁶, Joseph J. Fawcett², Dave Flowers⁶, Dea Fotopulos⁶, Tijana Glavina¹, Maria Gomez⁶, Eidelyn Gonzales⁶, David Goodstein¹, Lynne A. Goodwin², Deborah L. Grady², Igor Grigoriev¹, Matthew Groza³, Nancy Hammon¹, Trevor Hawkins¹, Lauren Haydu⁶, Carl E. Hildebrand², Wayne Huang¹, Sanjay Israni¹, Jamie Jett¹, Phillip B. Jewett², Kristen Kadner¹, Heather Kimball¹, Arthur Kobayashi^{1,3}, Marie-Claude Krawczyk², Tina Leyba², Jonathan L. Longmire², Frederick Lopez⁶, Yunian Lou¹, Steve Lowry¹, Thom Ludeman², Chitra F. Manohar³, Graham A. Mark², Kimberly L. McMurray², Linda J. Meincke², Jenna Morgan¹, Robert K. Moyzis², Mark O. Mundt², A. Christine Munk², Richard D. Nandkeshwar³, Sam Pitluck¹, Martin Pollard¹, Paul Predki¹, Beverly Parson-Quintana², Lucia Ramirez⁶, Sam Rash¹, James Retterer⁶, Darryl O. Rieke², Donna L. Robinson², Alex Rodriguez⁶, Asaf Salamov¹, Elizabeth H. Saunders², Duncan Scott¹, Timothy Shough², Raymond L. Stallings², Malinda Stalvey², Robert D. Sutherland², Roxanne Tapia², Judith G. Tesmer², Nina Thayer^{1,2}, Linda S. Thompson², Hope Tice¹, David C. Torney², Mary Tran-Gyamfi¹, Ming Tsai⁶, Levy E. Ulanovsky², Anna Ustaszewska¹, Nu Vo⁶, P. Scott White², Albert L. Williams², Patricia L. Wills², Jung-Rung Wu², Kevin Wu⁶, Joan Yang⁶, Pieter DeJong⁷, David Bruce², Norman A. Doggett², Larry Deaven², Jeremy Schmutz⁶, Jane Grimwood⁶, Paul Richardson¹, Daniel S. Rokhsar¹, Evan E. Eichler⁵, Paul Gilna², Susan M. Lucas¹, Richard M. Myers⁶, Edward M. Rubin^{1,4} & Len A. Pennacchio^{1,4}

¹DOE Joint Genome Institute, 2800 Mitchell Avenue, Walnut Creek, California 94598, USA

²Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

³Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA

⁴Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, California 94720, USA

⁵Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA

⁶Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, 975 California Avenue, Palo Alto, California 94304, USA

⁷Children's Hospital Oakland, Oakland, California 94609, USA

Human chromosome 16 features one of the highest levels of segmentally duplicated sequence among the human autosomes. We report here the 78,884,754 base pairs of finished chromosome 16 sequence, representing over 99.9% of its euchromatin. Manual annotation revealed 880 protein-coding genes confirmed by 1,670 aligned transcripts, 19 transfer RNA genes, 341 pseudogenes and three RNA pseudogenes. These genes include metallothionein, cadherin and iroquois gene families, as well as the disease genes for polycystic kidney disease and acute myelomonocytic leukaemia. Several large-scale structural polymorphisms spanning hundreds of kilobase pairs were identified and result in gene content differences among humans. Whereas the segmental duplications of chromosome 16 are enriched in the relatively gene-poor pericentromere of the p arm, some are involved in recent gene duplication and conversion events that are likely to have had an impact on the evolution of primates and human disease susceptibility.

The US Department of Energy (DOE) initiated the mapping and sequencing of human chromosome 16 in 1988 to contribute to the generation of a reference human genome sequence to be used in assessing the effects of radiation and for the study of human biology. This particular chromosome was in part targeted for sequencing because of the localization of the DNA repair gene *ERCC4* to the p arm of chromosome 16 (ref. 1), the availability of a unique flow-sorted chromosome-specific cosmid library², and access to a mouse-human hybrid cell panel enabling the localization of clones to discrete cytogenetic intervals³. Further interest in human chromosome 16 stemmed from the clustering of metallothionein genes on this chromosome, which participate in heavy metal transport and detoxification, coinciding with important biological interests of the DOE^{4,5}. Here we describe the finished human chromosome 16 sequence, which provides a reference for the further exploration of genomic sequence alterations and their relationship to human biology.

Mapping and sequencing

To provide the foundation for sequencing human chromosome 16, we constructed a physical map based on previous sequence-tagged site (STS) content maps⁶⁻⁸ with a minimal final tiling path of 716

clones, which include 618 bacterial artificial chromosomes (BACs), 79 cosmids, seven fosmids, five phage-derived artificial chromosomes (PACs), three yeast artificial chromosome (YAC) subclones, two P1 phages, two phage vectors and five genomic polymerase chain reaction (PCR) fragments. The final sequence contains four gaps, with two in each of the chromosome arms. One of the gaps is found in the highly duplicated pericentromeric region in the p arm, while two of the remaining non-pericentromeric gaps are resistant to stable cloning with conventional vectors, and efforts are ongoing to close the estimated ~25 kilobases (kb) of missing sequence using alternative vectors⁹. The final gap is found near the telomere of the q arm in a region of subtelomeric repeats distal to the last identifiable cosmid subclone (AC137934) of a 16q telomere half-YAC as previously described¹⁰.

The high degree of segmental duplication of chromosome 16, coupled with the multiple haplotypes represented in the numerous clone libraries comprising the tiling path, hindered efforts to construct a valid clone-based representation of this chromosome. To resolve this issue, we adopted a strategy of high depth clone coverage from a library constructed from a single individual¹¹. This enabled the determination of both of the diploid haplotypes across the segmentally duplicated intervals. Overall, these efforts resulted

in the generation of 78,884,754 base pairs (bp) of finished euchromatic sequence with an estimated accuracy¹² exceeding 99.9% and covering in excess of 99.9% of its euchromatin. Including the centromere and its adjacent heterochromatic portion of the q arm, sized together at 9.8 megabases (Mb) (see Methods), the total size of the chromosome is estimated at 88.7 Mb.

As a further assessment of the physical sequence, we compared it to the existing physical and genetic maps. We were able to account for all sequence-tagged sites from the Genethon¹³ microsatellite, the deCODE¹⁴ and the Marshfield¹⁵ genetic maps. We also compared the final DNA sequence with recombination distances in the deCODE female, male and sex-averaged meiotic maps (Fig. 1). We found the female recombination distances for chromosome 16 were similar to other human chromosomes, showing a linear relationship between recombination and physical distances at an average of 1.93 cM Mb⁻¹, excluding heterochromatin. However, the male meiotic map displayed substantial differences in the region from 17–72 Mb with a meiotic distance of only 22.5 cM, yielding an average of 0.50 cM Mb⁻¹. Finally, we found a marked increase in male recombination near the telomeres, exceeding 3 cM Mb⁻¹, consistent with other human chromosomes¹⁶.

Gene catalogue

We manually curated gene models as previously described¹⁷ and identified a total of 880 protein-coding gene loci (Table 1, Supplementary Information Table 1 and http://www.jgi.doe.gov/human_chr16) supported by 1,670 full-length (or nearly full-length) transcripts. These provided an average of 1.9 annotated transcripts per locus with 450 of the loci showing strong evidence for alternative splicing with two or more annotated messenger RNA transcripts. Additionally, 208 loci have 'expressed sequence tag' (EST) evidence for alternative splice forms, resulting in nearly 75% of loci displaying some evidence for alternative splice variants. Loci were further classified as 'known genes', 'novel genes' or 'pseudogenes', consistent with our previous definitions¹⁷, excluding loci without unique open reading frames, and *ab initio* predictions without supporting evidence. Seven hundred and seventy-one known genes were modelled on the basis of 2,435 Refseq transcripts as well as other complementary DNA sequence evidence in GenBank. Nearly one-third (36%) of these known genes were extended

by more than 50 bp at the 5' end and 18% at the 3' end relative to Refseq transcripts while maintaining their original open reading frame.

We identified thirty 'novel genes' based on cDNA sequence, spliced ESTs and protein similarity to known human or mouse genes, and we modelled an additional 79 putative novel genes using orthologous mouse cDNA sequences and *ab initio* predictions. Additionally, we annotated 19 tRNA genes and three tRNA pseudogenes based on previous data¹⁸. Finally we identified 341 pseudogenes and pseudogene fragments of which 120 appear to be non-processed because they displayed an exon structure similar to the parent locus and are therefore likely to have resulted from genomic duplication events. The remaining 221 appear to be processed pseudogenes, presumably resulting from viral retrotransposition of spliced mRNAs or from mitochondrial genome insertion. At least one frameshift or premature stop codon (in comparison to the parent gene) was identified in 233 pseudogenes and the remaining 108 were processed pseudogenes lacking introns and displaying poly-A's in the adjacent genomic sequence. This supports the likely nonfunctional nature of these vestigial genes. To assess the quality of our pseudogene collection, we compared it to an earlier analysis¹⁹ describing 250 processed pseudogenes on chromosome 16. Initially we were able to map 233 of these 250 pseudogenes to 429 loci on chromosome 16 using BLAT²⁰ with 100% coverage and >99% identity. We then eliminated loci consisting of repetitive DNA²¹ (Smit, A. F. A. and Green, P., unpublished results), those covering less than 50% of the parent gene and cases where there was clearly a retained intron/exon structure. This resulted in 146 processed pseudogenes in agreement between a previous study¹⁹ and our study, and suggested that our manual curation of the finished sequence identified 75 additional members.

Large structural polymorphisms

We observed several large structural polymorphisms based on the finished sequence of chromosome 16, which were often associated with segmental duplications. For instance, we further characterized a previously described stable length polymorphism within the 16p subtelomeric region^{22,23}. Whereas the shortest and most common allele was previously finished (represented in NCBI Build 35), we isolated and sequenced the majority of a longer allele derived from a 16p telomere half-YAC, located within close proximity of the TTAGGG telomere repeat as defined in ref. 10. This allele is ~137.5 kb longer than the current assembly, however the shorter allele is not simply a truncation of the longer form; rather the telomeric 21,056 bp of the short allele is not present in the long allele

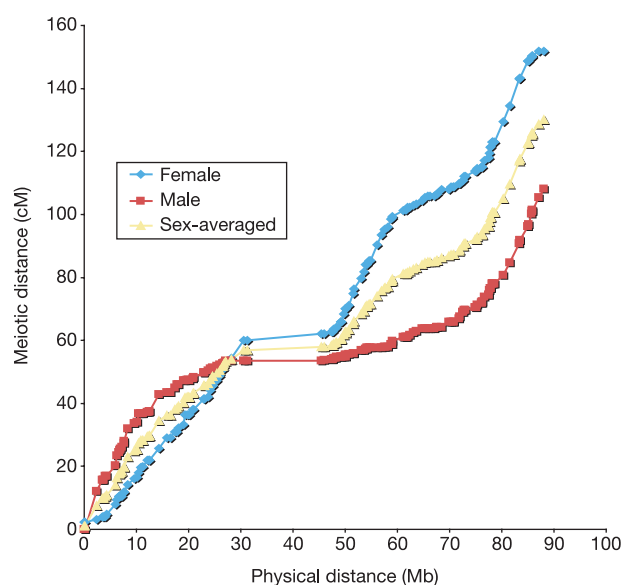


Figure 1 Comparison of meiotic distance to the physical map of chromosome 16, from the telomere of the short arm to the telomere of the long arm and reading left to right.

Table 1 Chromosome 16 sequence features

Gap-free size (finished bp)	78,884,754
Protein coding genes	880
Processed pseudogenes	221
Non-processed pseudogenes	120
Protein-coding genes per Mb	11.2
Average percentage of GC content	44.7
Protein-coding transcripts	1,670
Average transcripts per gene	1.9
Percentage Alu coverage	16.4
Percentage L1 coverage	11.8
Percentage L2 coverage	2.6
Total percentage repeat masked	47.8
(Ensembl PCG)	960
(Ensembl PCT)	1,441
Ensembl genes per Mb	12.16
Ensembl transcripts per gene	1.5
Human/rodent CNSs	5,187
CNSs per gene	5.9
CNSs per Mb	65.8
Human/mouse/dog/chicken CNSs	1,862
CNSs per gene	2.1
CNSs per Mb	23.6

PCG, protein-coding genes; PCT, protein-coding transcripts; CNS, conserved noncoding sequence; L1, long interspersed nucleotide element 1; L2, long interspersed nucleotide element 2.

and the telomeric 158,607 bp of the long allele is not shared with the short allele. Both of these unique regions contain genes with the short allele containing a putative gene(s) represented by cDNAs MGC:75272 and MGC:52000 and with the long allele containing genes encoding hypothetical protein XP_375548 (similar to septin), hypothetical protein XP_379920 (similar to capicua) and beta-tubulin 4Q (AAL32434).

We also identified one of the most extensively duplicated regions on chromosome 16 corresponding to a 500-kb interval at 16p11.2-12.1 composed of approximately 54 intrachromosomal duplications (Fig. 2 and Supplementary Table 2). This interval includes seven full or partial gene duplicates including the eukaryotic translation initiation factor 3 subunit 8 (*EIF3S8*), sulphotransferase 1A (*SULT1A1*) and the Batten disease gene (*CLN3*). Assembly of the region was initially complicated by the fact that the duplications were long (~200 kb) and showed an extraordinary degree of homology (98.33%). During the mapping of this region, sequence for a second haplotype variant from the RPCI-11 BAC library was nearly completed except for one gap of ~100 kb. Sequence

comparison of these two haplotypes (EIFvar1 and EIFvar2) revealed a 452-kb inversion between them (Fig. 2). Analysis of the break-points suggests that a large duplication palindrome is responsible for this rearrangement.

Finished sequence was also generated across a recently duplicated 360-kb polymorphism of the human homologue of the hydrocephalus inducing gene (*HYDIN*) at 16q22, which is inserted in some humans at chromosome 1q21.1. The RPCI-11 BAC library seems to be heterozygous for this insertional polymorphism, with the current genomic assembly for chromosome 1 containing the haplotype version lacking the insertion. We further investigated a recently described²⁴ copy number polymorphism between 16p11.2 and 6p25, which contains the *DUSP22* gene. On the basis of extensive drafting of RPCI-11 BACs in the region and comparisons with drafted clones from monochromosomal libraries for chromosomes 6 and 16, we were able to determine that the RPCI-11 library is homozygous and lacking the *DUSP22* duplication on chromosome 16. Taken together, these recently arisen large structural polymorphisms are striking examples of variability in the human

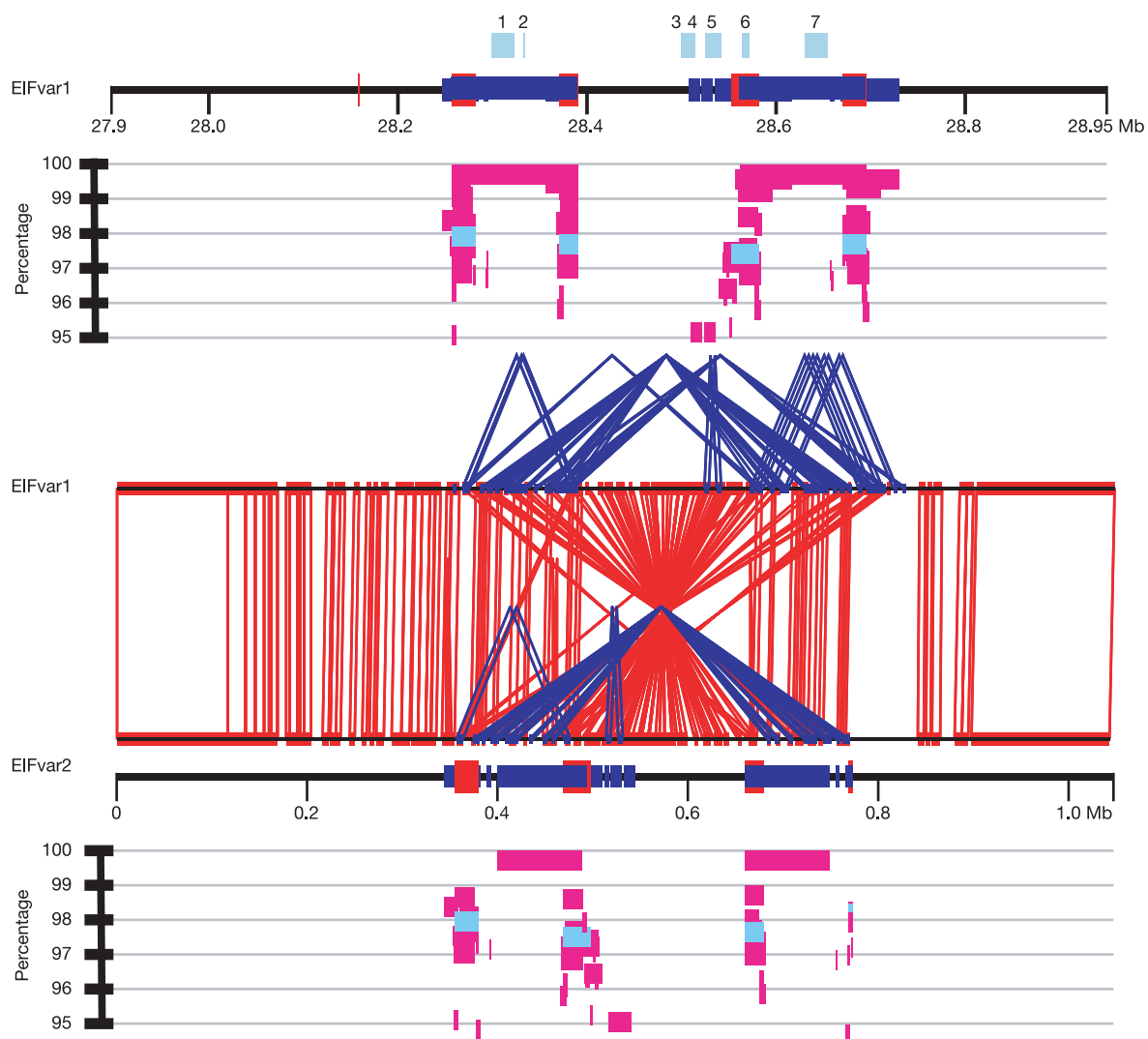


Figure 2 A 450-kb inversion haplotype on chromosome 16. The duplication and inverted structure for two chromosome 16 haplotypes (EIFvar1 and EIFvar2) are compared. Top and bottom panels: interchromosomal (red) and intrachromosomal duplications (blue) alignments (>90%, >1 kb) are depicted as a function of percentage identity below the horizontal line with different colours corresponding to the location of the pairwise alignment on different human chromosomes (chromosome 16 is shown as magenta, chromosome 18 as sky blue). The middle panel shows a 450-kb inversion between

EIFvar1 and EIFvar2, using Miropeats (threshold = 3,000). Interhaplotype (red) and intrahaplotype (blue) sequence alignments are shown based on chromosome assembly for EIFvar1. A palindromic duplication structure (200 kb) demarcates the breakpoint region. Genes are depicted as light-blue bars above the horizontal line in the top panel. These include: (1) eukaryotic translation initiation factor 3, subunit 8 (*EIF3S8*); (2) LOC39068; (3) LOC11286; (4) sulphotransferase 1A (*SULT1A2*); (5) sulphotransferase 1A (*SULT1A1*); (6) JGI-495; and (7) *EIF3S8*.

genome and support a potential mechanism that contributes to phenotypic or disease susceptibility differences among humans. It is worth noting that 91 genes on chromosome 16 are located within segmental duplications, any of which could be unstable and challenge researchers studying phenotypes linked to these gene-containing regions. These observations are particularly relevant on the basis of recent findings^{24,25} of abundant copy number polymorphisms within the genomes of normal individuals, which include those described here.

Duplication analysis of chromosome 16

We performed a detailed analysis of duplicated genomic sequence ($\geq 90\%$ sequence identity and ≥ 1 kb in length) comparing chromosome 16 against the July 2003 assembly of the human genome. We found that 9.89% (7.8 Mb) of chromosome 16 consists of segmental duplications (Supplementary Table 2). In comparison to other finished chromosomes and to the human genomic average (5.3%), chromosome 16 is one of the most enriched chromosomes for segmental duplications (Supplementary Table 2 and Supplementary Fig. 1). Nearly 9% of genome-wide human duplication alignments map to this chromosome. Intrachromosomal duplications are longer and show higher sequence identity when compared with interchromosomal duplications (Fig. 3a and Supplementary Fig. 2). Whereas there is a general inverse correlation between duplication length and divergence, the effect is most pronounced for intrachromosomal duplication in which the average length of duplicated DNA exceeds 16 kb. A clear bimodal distribution pattern of sequence identity is distinguishable based on the distribution pattern of the alignments. Most interchromosomal duplication alignments show 93–95% sequence identity whereas intrachromosomal duplications show greater than 97% sequence identity, consistent with a recent expansion of intrachromosomal duplications along the chromosome^{26,27}. On the basis of substitution rates between great apes, we estimate that as much as 7% of the mass of human chromosome 16 was added by segmental duplication events within the last 10 million years of human evolution²⁸.

Segmental duplications are particularly clustered along the p arm of the chromosome (Supplementary Figs 1 and 3). As described previously²⁹, the 16p11 pericentromeric region represents the largest zone of interchromosomal duplications (Fig. 3b) accounting for 44% (937 of 2,146) of the total number of chromosome 16 alignments (Supplementary Table 4) and 55% (752 of 1,365) of all chromosome 16 interchromosomal alignments. Most of the interchromosomal duplications in this region map to the pericentromeric regions of other chromosomes (Fig. 3b). Large tracts of interstitial alpha-satellite DNA have been finished within proximal

16p11 and it is possible that such sequences have played a part in the frequent evolutionary exchange of pericentromeric DNA among non-homologous chromosomes³⁰. In stark contrast to 16p11, there is little evidence for extensive pericentromeric duplication on the q arm despite the fact that centromeric satellite boundary sequences have been traversed.

An additional 19 blocks of extensive duplication (>100 kb and >5 duplication alignments) were identified within the euchromatic portion of chromosome 16. These regions are composed of as many as 119 underlying duplicons (also known as low-copy repeats on chromosome 16, LCR16) that have been juxtaposed in different combinations within the duplication blocks. These contain various genes and gene fragments, such as *NPIP*, *SULT1A*, *EIF3S8* and *SMG1* (Supplementary Table 3). Most are duplicated several times in varying copy numbers with a high degree of sequence identity to their putative ancestral genes. Most seem to have been duplicated in concert with LCR16a, a segment that contains one of the most rapidly evolving gene families of the human genome^{27,31}.

Comparative genomics

We compared human chromosome 16 to the chimpanzee, dog, mouse³², rat³³, chicken and fish³⁴ (*Fugu rubripes*) draft genomes to further explore the evolution and constraint of sequences found along this chromosome. By first building segmental maps from DNA alignments of all the vertebrate species described above, we were able to examine the global homologous chromosomal relationships between these vertebrate genomes and human chromosome 16 (see Methods). We found no major rearrangements relative to the homologous chimpanzee chromosome 18. Comparison with the mouse and rat genomes revealed 26 chromosomal segments unbroken in any of the three species, ranging in size from 250 kb to 10.7 Mb (Fig. 4a). Further addition of the chicken genome to the multi-dimensional map yielded 33 segments ranging in size from 250 kb to 8.7 Mb (Fig. 4a). These segmental maps provide the substrates to precisely define the breakpoints that, in some cases, may have disrupted gene loci in the species containing the rearrangement.

We next identified slowly evolving regions, presumably under evolutionary constraint, through fine-scale DNA comparison of chromosome 16 with other vertebrate genome assemblies. Four different species combinations were selected to represent the accessible range of vertebrate evolutionary divergence times: human/mouse/rat, human/mouse/rat/dog, human/mouse/dog/chicken, and human/mouse/*Fugu* (see Methods). To explore potential noncoding functional elements on chromosome 16, the results were filtered for overlap with annotated genes, spliced ESTs or mRNAs in human, mouse and rat, which resulted in the

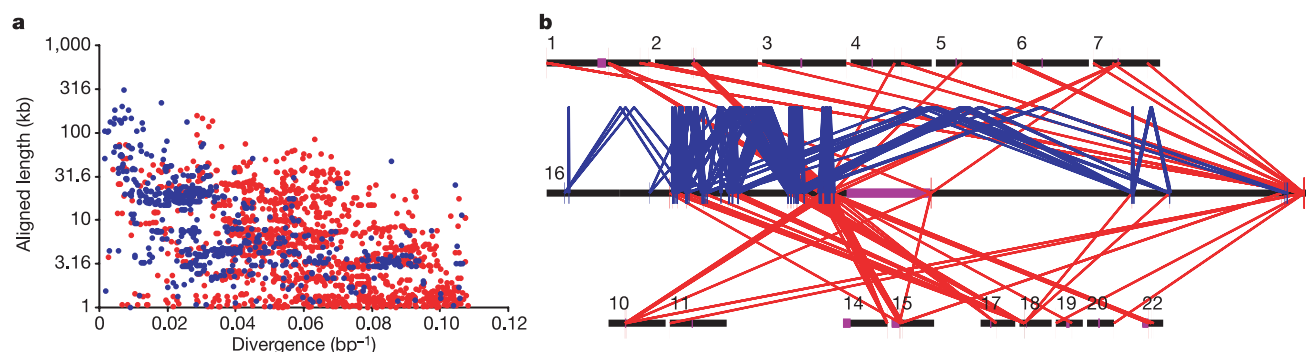


Figure 3 Chromosome 16 segmental duplications. **a**, The scatter plot depicts the length (log 10) and divergence of inter- (red) and intra- (blue) chromosomal segmental duplication. Divergence is calculated as the number of substitutions per site between the two sequences. **b**, The parasight view depicts the pattern of interchromosomal (red) and

intrachromosomal (blue) duplications (>20 kb, $>95\%$) for chromosome 16. Chromosome 16 is drawn at $20\times$ greater scale than the other chromosomes. Centromeres are shown as purple bars.



Figure 4 Comparative analysis of human chromosome 16. **a**, Segmental homology maps between human chromosome 16 and the chimpanzee, mouse, rat, dog and chicken genomes. Syntenic segments are colour-coded by chromosome, with arrowheads indicating the strand. **b**, Gene density (blue) and noncoding conservation density (magenta) over the entire chromosome. Densities are normalized so that the darkest shade in each track denotes 3.5 times the genomic average. **c**, Conservation in the second largest human/mouse/dog/chicken synteny block on human chromosome 16, which spans 8.19 Mb at 16q21 (*NCBI/34* chr16: 58,626,110–66,811,606), and contains

four cadherin genes. The upper plot shows coding (blue) and noncoding (magenta) conservation *P*-values in the human/mouse/rat comparison. The lower plot shows the human/mouse/dog/chicken comparison. **d**, Similar plots of ENCODE region ENr313 (*NCBI/34* chr16: 62,051,662–62,551,661), which lies near the centre of the gene-poor region in **c**. **e**, ENCODE region ENr211 (*NCBI/34*-chr16:25,839,478–26,339,477), another gene-poor region on 16p12.1. Rat is excluded because of a large sequencing gap. In **c**, **d** and **e**, the height of the bars is proportional to $-\log(\text{conservation } P\text{-value})$ (Gumby and Rank-VISTA, see Methods).

identification of 5,187 discrete conserved noncoding regions between human/mouse/rat, 6,159 between human/mouse/rat/dog, 1,862 between human/mouse/dog/chicken, and 191 between human/mouse/*Fugu* (Fig. 4b and Supplementary Table 1). Compared with genome-wide averages, the densities of human/mouse/rat and human/mouse/dog/chicken elements were only slightly higher for human chromosome 16 (Supplementary Table 1). In contrast, human/mouse/*Fugu* elements are present at ~2.4 times the genome-wide density, indicating that although chromosome 16 as a whole has had 'normal' levels of noncoding constraint since the mammal/bird split, it has conserved more ancient functions to a surprising degree. Functional studies on these conserved elements are warranted to assess their possible biological activity in the ~98% of the human genome that is noncoding.

We further explored an 8.7-Mb region at 16q12, on the basis of extreme features of evolutionary conservation. This region was first identified as the largest unbroken synteny segment between human/mouse/dog/chicken on chromosome 16 and contains 59% (112 of 191) of the human/mouse/*Fugu* noncoding elements. These elements are entirely clustered in a gene-poor 5-Mb subregion, which contains at least six developmental transcription factors, including *SALL1* and three iroquois genes (*IRX3*, *IRX5* and *IRX6*). This clustering is an example of the general bias of human–fish conserved sequences towards developmental genes³⁵. Interestingly, at least nine of these human/mouse/*Fugu* elements have significant sequence similarity to counterparts in the paralogous *IRX* gene cluster on chromosome 5, which is similarly located in a 'forest' of human–fish conservation³⁶. *In vivo* mouse transgenic data indicate that a significant percentage of these *IRX* conserved noncoding sequences behave as gene enhancers (Pennacchio, L. A., unpublished observation), suggesting that in addition to the well described conservation of the protein-encoding portions of genomic duplications, evolutionary constraint is also observable in adjacent gene regulatory sequences following genomic duplication events. This synteny block is an outlier even in terms of more recent noncoding conservation, with 917 (105 per Mb) human/mouse/rat and 590 (67.5 per Mb) human/mouse/dog/chicken elements.

The second longest chromosome 16 synteny block in human/mouse/dog/chicken neighbours the highly conserved *SALL1-IRX* segment and is similar in length (8.19 Mb) (Fig. 4c). Once again this region is gene poor, with its telomeric 7.6 Mb containing only three annotated genes, all members of the cadherin family: *CDH8*, *CDH11* and *CDH5*. Within the full 8.19-Mb interval, we identified 968 (118 per Mb) human/mouse/rat conserved noncoding sequences. This is twice the genome-wide density, as was the case in the *SALL1-IRX* region. However, in stark contrast to the neighbouring *SALL1-IRX* region, this synteny block has no noncoding conservation between human/mouse/*Fugu*, suggesting that its noncoding functions, though just as constrained among mammals, are more diverged in distant species.

As a special category of constrained DNA, we also searched for ultra-conserved noncoding sequences, recently defined by the stringent criterion of at least 200 bp in length and 100% identity between the human, mouse and rat genomes³⁷. Of the 482 ultra-conserved elements found in the entire human genome, 15 (3.1%) were found on chromosome 16, with 11 having some evidence of being transcribed and processed into mature mRNAs. The above-mentioned bias towards developmental genes has also been noted³⁷ for ultra-conserved human/rodent elements. Indeed, 9 of the 15 ultra-conserved elements found on chromosome 16 lie in the same *SALL1-IRX* synteny block that contains the mammal/fish conservation cluster. This contrasts with the similarly sized cadherin synteny block that contains no human–fish noncoding conservation and only one ultra-conserved element.

Finally, three regions on chromosome 16 have been selected by the National Human Genome Research Institute as part of the Encyclopaedia of DNA Elements (ENCODE) project, an effort

aimed at rigorously analysing 1% of the human genome sequence³⁸ (<http://www.genome.gov/10005107>). These three ENCODE regions include the well-studied alpha-globin-containing interval (ENm008) and two randomly chosen regions (ENr211 on 16p12.1 and ENr313 on 16q21). Interestingly, ENr313 is located within the large cadherin gene desert described above and is completely devoid of genes (Fig. 4d). Nonetheless, it harbours the same high density of human/mouse/rat and human/mouse/dog/chicken conserved noncoding elements as the rest of the cadherin synteny block, suggesting the presence of numerous unassigned functional sequences within this region. Ongoing studies by ENCODE will better define the overlap of functionality and comparative sequence data such as that presented here.

Discussion

The primary sequence of human chromosome 16, as well as the human genome as a whole, now provides a key foundation for ongoing efforts such as ENCODE to deeply annotate all types of information encoded in our genome. This represents an enormous long-term challenge because genomic signatures embedded within the sequence of DNA perform a vast number of different operations across the trillions of cells within our bodies. These features range from relatively easily identified genes, to sequences involved in gene regulation—which use a plethora of signals to determine when and where a given gene is expressed and under what conditions—to probably even more complicated features such as higher-order chromosome structure and DNA involvement in replication and repair. It is inspiring to reminisce that it was only 50 years ago that we had our first glimpse into the structure of DNA, which provided the foundation for generating the nearly entire human euchromatic sequence. The next 50 years will probably also bring similarly impressive gains and enable us to precisely relate our primary genomic sequence to functional genomic signatures and their relationship to human biology. □

Methods

Sizing of heterochromatic gaps

To estimate the size of the alpha satellite bands (16p11.1–16q11.1) encompassing the centromere and the satellite II heterochromatin in band 16q11.2, we used contour-clamped homogeneous electric field (CHEF) pulsed-field gel electrophoresis at various pulse times to resolve macrorestriction fragments between 100 kb and >7,000 kb. DNA from CY18 (a mouse–human hybrid containing a single human chromosome 16) was digested with several different rare cutting restriction enzymes and separated on CHEF gels. Hybridization to blots of these gels with 16-1 (16-specific alpha satellite) and pHuR 195 (16-specific satellite II) probes revealed a single band of alpha satellite (in three different enzyme digests) that did not overlap with any satellite II bands (data not shown). The smallest of these bands was an 1,800-kb Xho I fragment, which provided an upper size limit for the alpha satellite array, encompassing the centromere on chromosome 16. *Sall* fragmented the satellite II heterochromatin into well resolved large restriction fragments without cutting within the alpha satellite array. The sum of the *Sall* satellite II fragments was estimated at ~7,800 kb providing an upper size limit of the 16q11.2 satellite II heterochromatin at nominally 8 Mb. Together these account for 9.8 Mb of unsequenced heterochromatin encompassing cytogenetic bands 16p11.1–16q11.2, although it is likely that we did sequence partially into the boundaries of these regions in the adjacent tiling set clones.

Segmental duplication analysis

We used a BLAST-based detection scheme³⁹ to identify all pairwise similarities representing duplicated regions (≥ 1 kb and $\geq 90\%$ identity) within the finished sequence of chromosome 16 and compared it with all other chromosomes in the NCBI genome assembly (build 34). A total of 2,146 pairwise alignments representing 26.12 Mb of aligned basepairs and 7.8 Mb of non-redundant duplicated bases were analysed on chromosome 16. The program Parasight (<http://humanparalogy.gene.cwru.edu/parasight/>) was used to generate images of pairwise alignments. Divergence of duplication, the number of substitutions per site between the two sequences, were calculated using Kimura's two-parameter method, which corrects for multiple events and transversion/transition mutational biases⁴⁰. Analysis of haplotype structural variation was performed using the program Miropeats (threshold = 3,000)⁴¹. Gene content of each 1% duplicated regions of 90–100% identity was analysed using a non-redundant/non-overlapping set of known genes. A gene feature (exon) was considered duplicated if >50 bp of the feature overlapped duplication. Thus, exons less than 50 bp were lost in this analysis.

Pseudogene identification

Pseudogenes were defined as gene models built by homology to known human genes in which the alignment between the model and the homologue shows at least one stop codon or frameshift. We identified homologues⁴² of human IPI (International Protein Index; <http://www.ebi.ac.uk/IPI/IPIhelp.html>) proteins on repeat-masked²¹ (Smit, A. F. A. and Green, P., unpublished results) genomic chromosome 16 sequence. For each such fragment of genomic sequence we built gene models using the GeneWise⁴³ program. Overlapping models were then clustered and the top-scoring model was analysed for the presence of premature stop codons and frameshifts. Remaining models were then manually checked to confirm their pseudogene status.

Comparative analysis

Multi-species segmental homology maps were computed using PARAGON (v2.2; Couronne, unpublished work), which is based on BLASTZ⁴⁴ pairwise alignments of all genomes to human. After filtering out segments shorter than 250 kb in humans, MLAGAN⁴⁵ alignments of homologous blocks were scanned for evolutionarily conserved regions using Gummy (v1.5; Prabhakar, unpublished work). These were visualized using Rank-VISTA (Prabhakar, unpublished work). Gummy goes through the following three-step process to identify statistically significant conservation in the input global alignment: (1) first, noncoding regions in the alignment are used to estimate the local neutral mutation rates⁴⁶ between all pairs of aligned sequences. The rates are used to derive a log-likelihood scoring scheme for slow versus neutral evolution⁴⁷, in which the slow rate is set at half the neutral rate; (2) each alignment position is then assigned a conservation score using a phylogenetically weighted sum-of-pairs scheme; (3) finally, a dynamic programming step scans the alignment for high-scoring segments (conserved regions) of any length. Conserved regions detected in this manner are assigned *P*-values using the same statistical formalism⁴⁸ as the BLAST algorithm⁴². Whereas BLAST assigns *P*-values relative to random permutations of the query and target sequences, Gummy *P*-values relate to random permutations of the columns in the input alignment. Here, all the results were generated using a Gummy *P*-value threshold of 0.01 and a baseline human sequence length of 100 kb. Conserved noncoding regions were defined as conserved segments that overlap annotated exons, spliced ESTs or mRNAs from human, mouse or rat over no more than 25% of their length. At a Gummy *P*-value threshold of 0.01, 2.2% of the ungapped positions in the human genome were assigned to human/mouse/rat conserved noncoding segments.

Received 9 September; accepted 15 November 2004; doi:10.1038/nature03187.

- Siciliano, M. J. Chromosomal assignment of human genes coding for DNA repair functions. *Isozymes Curr. Top. Biol. Med. Res.* **15**, 217–223 (1987).
- Deaven, L. L. *et al.* Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 159–167 (1986).
- Callen, D. F. *et al.* High-resolution cytogenetic-based physical map of human chromosome 16. *Genomics* **13**, 1178–1185 (1992).
- Hildebrand, C. E. & Enger, M. D. Regulation of Cd2+/Zn2+-stimulated metallothionein synthesis during induction, deinduction, and superinduction. *Biochemistry* **19**, 5850–5857 (1980).
- Stallings, R. L., Munk, A. C., Longmire, J. L., Hildebrand, C. E. & Crawford, B. D. Assignment of genes encoding metallothioneins I and II to Chinese hamster chromosome 3: evidence for the role of chromosome rearrangement in gene amplification. *Mol. Cell. Biol.* **4**, 2932–2936 (1984).
- Han, C. S. *et al.* Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. *Genome Res.* **10**, 714–721 (2000).
- Doggett, N. A. *et al.* An integrated physical map of human chromosome 16. *Nature* **377**, 335–365 (1995).
- Cao, Y. *et al.* A 12-Mb complete coverage BAC contig map in human chromosome 16p13.1-p11.2. *Genome Res.* **9**, 763–774 (1999).
- Kouprina, N. *et al.* Construction of human chromosome 16- and 5-specific circular YAC/BAC libraries by *in vivo* recombination in yeast (TAR cloning). *Genomics* **53**, 21–28 (1998).
- Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
- Osoegawa, K. *et al.* A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* **11**, 483–496 (2001).
- Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365–368 (2004).
- Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
- Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
- Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
- Yu, A. *et al.* Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953 (2001).
- Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
- Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
- Zhang, Z., Harrison, P. M., Liu, Y. & Gerstein, M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* **13**, 2541–2558 (2003).

- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000).
- Flint, J. *et al.* The relationship between chromosome structure and function at a human telomeric region. *Nature Genet.* **15**, 252–257 (1997).
- Wilkie, A. O. *et al.* Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64**, 595–606 (1991).
- Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- Iafate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
- Loftus, B. *et al.* Genome duplications and other features in 12 Mbp of DNA sequence from human chromosome 16p and 16q. *Genomics* **60**, 295–308 (1999).
- Johnson, M. E. *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514–519 (2001).
- Chen, F. C. & Li, W. H. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
- She, X. *et al.* The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857–864 (2004).
- Guy, J. *et al.* Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10p. *Genome Res.* **13**, 159–172 (2003).
- Eichler, E. E. *et al.* Divergent origins and concerted expansion of two segmental duplications on chromosome 16. *J. Hered.* **92**, 462–468 (2001).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
- Boffelli, D., Nobrega, M. A. & Rubin, E. M. Comparative genomics at the vertebrate extremes. *Nature Rev. Genet.* **5**, 456–465 (2004).
- Schmutz, J. *et al.* The DNA sequence and comparative analysis of human chromosome 5. *Nature* **431**, 268–274 (2004).
- Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
- The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
- Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
- Parsons, J. D. Miropeats: graphical DNA sequence comparisons. *Comput. Appl. Biosci.* **11**, 615–619 (1995).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
- Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
- Bruno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).
- Cooper, G. M. *et al.* Characterization of evolutionary rates and constraints in three mammalian genomes. *Genome Res.* **14**, 539–548 (2004).
- Boffelli, D. *et al.* Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391–1394 (2003).
- Karlin, S. & Dembo, A. Limit distributions of maximal segmental score among Markov-dependent partial sums. *Adv. Appl. Prob.* **24**, 113–140 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the International Chimpanzee Sequencing Consortium for pre-publication access to and permission to analyse the relevant portions of the chimpanzee genomic sequence; the Broad Institute for pre-publication access to the dog genome assembly; and the Washington University Genome Sequencing Center for pre-publication access to the chicken genomic assembly. We also thank D. Gordon of the University of Washington for his assistance in developing and customizing finishing tools, and T. Furey and G. Schuler for their efforts towards assessing the quality and completeness of our assemblies. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory, Lawrence Berkeley National Laboratory, Los Alamos National Laboratory, and Stanford University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.M.R. (emrubin@lbl.gov) or L.A.P. (LAPennacchio@lbl.gov). The sequence described here has been deposited in public databases with accession number NC_000016.

Transposition of *hAT* elements links transposable elements and V(D)J recombination

Liqin Zhou¹, Rupak Mitra¹, Peter W. Atkinson², Alison Burgess Hickman³, Fred Dyda³ & Nancy L. Craig¹

¹Howard Hughes Medical Institute and Department of Molecular Biology & Genetics, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA

²Department of Entomology and Institute for Integrative Genome Biology, University of California, Riverside, California 92521, USA

³Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, USA

Transposons are DNA sequences that encode functions that promote their movement to new locations in the genome. If unregulated, such movement could potentially insert additional DNA into genes, thereby disrupting gene expression and compromising an organism's viability. Transposable elements are classified by their transposition mechanisms and by the transposases that mediate their movement. The mechanism of movement of the eukaryotic *hAT* superfamily elements was previously unknown, but the divergent sequence of *hAT* transposases from other elements suggested that these elements might use a distinct mechanism. Here we have analysed transposition of the insect *hAT* element *Hermes* *in vitro*. Like other transposons, *Hermes* excises from DNA via double-strand breaks between the donor-site DNA and the transposon ends, and the newly exposed transposon ends join to the target DNA. Interestingly, the ends of the donor double-strand breaks form hairpin intermediates, as observed during V(D)J recombination, the process which underlies the combinatorial formation of antigen receptor genes. Significant similarities exist in the catalytic amino acids of *Hermes* transposase, the V(D)J recombinase RAG, and retroviral integrase superfamily transposases, thereby linking the movement of transposable elements and V(D)J recombination.

Transposons are discrete mobile DNA segments that have been found in virtually every genome examined. The major fraction of some genomes, including the human genome, is composed of transposable elements^{1,2}. The movement of many transposable elements involves a DNA intermediate whose ends are specifically recognized by the element-encoded transposase. These ends are processed by transposase-mediated DNA breakage reactions that expose the element termini and join them to the target DNA^{3,4}. Such elements include both DNA-only elements whose termini are short terminal inverted repeats (TIRs), represented by the prokaryotic IS50/IS10, transposon Tn7, and Mu elements as well as the eukaryotic P and Tc1/Mariner elements, and elements such as retroviruses that generate the integrating DNA version of their genome by reverse transcription of an RNA intermediate. Almost all of the elements with DNA intermediates whose transposition has been investigated biochemically belong to the retroviral integrase superfamily^{3,5}.

Another major family of transposons is the widespread eukaryotic *hAT* superfamily that includes active elements in fungi, plants and animals, including vertebrates^{6–8}; these elements have not yet been studied at the molecular level. *hAT* elements^{6,9} include *hobo* of *Drosophila* and the closely related *Hermes* element first identified in housefly¹⁰, McClintock's *Ac* element of maize, and the snapdragon *Tam3* element. They share a number of amino-acid motifs distinct from those shared among the retroviral integrase superfamily⁶. Using purified *Hermes* transposase, we have determined the mechanism of DNA breakage and joining that underlies *hAT* element transposition and have identified amino acids critical for the catalytic steps of recombination.

Hermes transposase binds the transposon ends

The extreme ends of *Hermes* are imperfect TIRs 17 base pairs (bp) long. They are likely to be central to recombination, but other sequences in the ends that may be required for transposition remain to be defined^{10,11}. DNA band shift assays reveal that purified *Hermes* transposase produced in *Escherichia coli* binds specifically

to a 131-bp fragment containing 80 bp of the *Hermes* left end (*Hermes*-L) flanked by housefly genomic DNA¹⁰ and to a 178-bp fragment containing 110 bp of the *Hermes* right end (*Hermes*-R); much less binding is observed with *Hermes*-R (Fig. 1). The many species observed with *Hermes*-L probably represent the binding of several transposase protomers because *Hermes* transposase binds

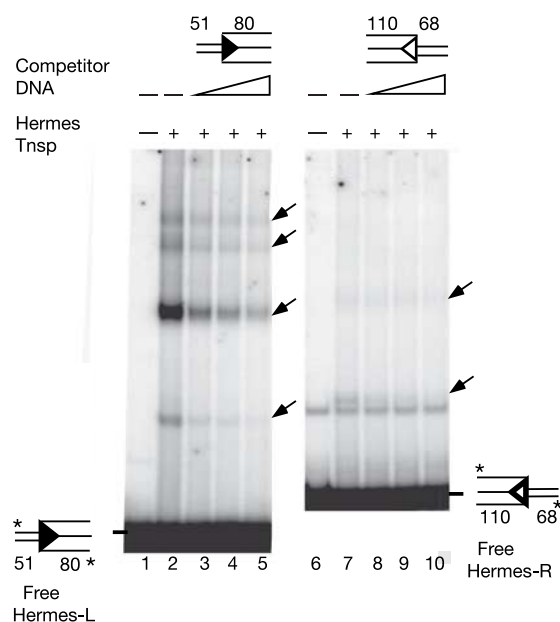


Figure 1 *Hermes* transposase binds specifically to the ends of *Hermes*. DNA binding assays using *Hermes* transposase (Tnsp) and radiolabelled *Hermes*-L (lanes 1–5) and *Hermes*-R (lanes 6–10) end fragments are shown; positions of radiolabelling are indicated by asterisks. Specific competitor DNA is unlabelled. Arrows indicate nucleoprotein complexes. The convention used throughout the figures is that the filled triangle indicates the TIR of *Hermes*-L and the open triangle the TIR of *Hermes*-R.

multiple positions on this DNA segment by DNA footprinting assays (J. Spencer, P. Kuduvali, X. Li and N.L.C.; R.H. Hice, T.A. Laver and P.W.A., unpublished observations). Some of the *Hermes*-L complexes might also reflect pairing between end segments. The presence of multiple transposase binding sites in a single end and protein bridging of transposon ends has been observed in many other systems⁴.

Hermes transposase makes hairpin ends

Genetic analysis suggests that *Hermes* transposes by excision from a donor site, followed by insertion into a target site¹¹. To probe these steps, we incubated *Hermes* transposase with a 194-bp fragment containing 91 bp of *Hermes*-L flanked by housefly genomic DNA and a target plasmid, then examined the resulting transposition intermediates and products (Fig. 2). *Hermes*-mediated

DNA-processing reactions require Mn^{2+} or Mg^{2+} ; more recombination is often seen with Mn^{2+} (see below and Supplementary Information) and the sequence selectivity of some DNA cleavages is greater with Mn^{2+} than Mg^{2+} (Supplementary Fig. 1). Cleavage of the flanking donor-site DNA from *Hermes*-L via a double-strand break results in two products on a native gel (Fig. 2a). The slower-migrating flanking DNA fragment increases in amount as the incubation time is extended; however, little of the faster-migrating *Hermes*-L segment accumulates, because it joins to the target plasmid forming a product which is not resolved on this gel (see Fig. 3). Cleavage of a similar *Hermes*-R fragment was also observed but at a much lower level (data not shown).

On a denaturing gel, two reaction products are also evident (Fig. 2b): the faster-migrating species of about 90 nucleotides (nt) is the top strand of *Hermes*-L, resulting from a single-strand break that

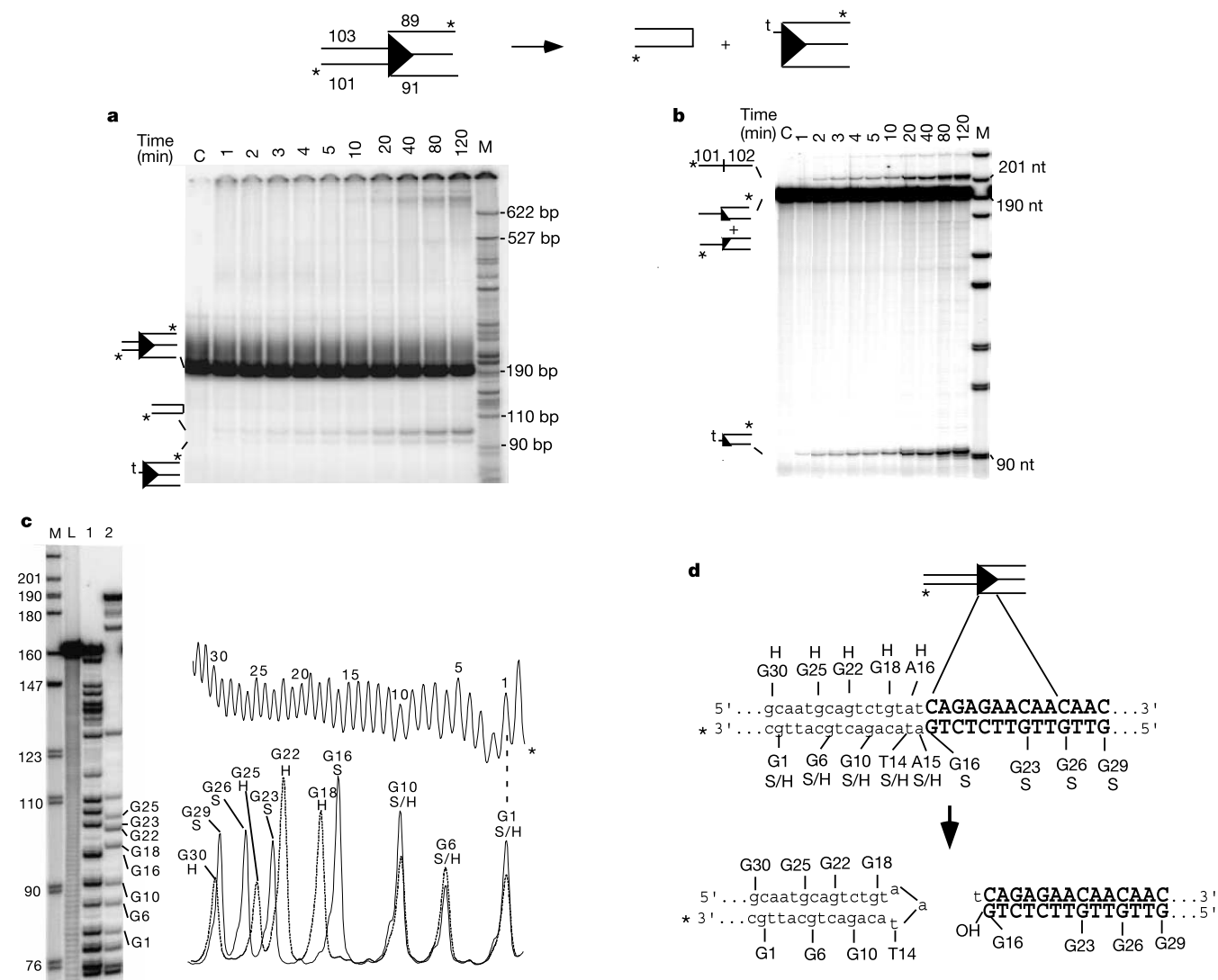


Figure 2 *Hermes* transposition involves hairpin formation on the flanking donor-site DNA. Transposition reactions containing Mn^{2+} were performed with *Hermes* Tnp and a DNA fragment containing *Hermes*-L flanked by housefly genomic DNA. The schematic diagram represents the *Hermes* double-strand break reaction. The 't' on the 5' transposon end in **a** and **b** indicates the nucleotide of flanking donor-site DNA that remains attached to that end. **a**, Reaction products on a native polyacrylamide gel reveal that the double-strand break reaction separates the transposon end from the flanking donor-site DNA. Higher-molecular-mass strand-transfer products that are not resolved in this gel are trapped in the wells (see Fig. 3). **b**, Reaction products on a denaturing polyacrylamide gel reveal that

the flanking donor-site DNA forms a hairpin structure with a lower mobility than the substrate. Lane C, buffer control; lane M, radiolabelled *Msp*I-digested pBR322 DNA. **c**, Analysis of the apparent higher-molecular-mass product by Maxam–Gilbert G sequencing establishes its hairpin nature. Lane L, substrate DNA treated with hydroxy radicals to provide length markers; lane 1, substrate DNA; lane 2, hairpin product. Positions in the substrate strand are designated 'S' and those in the hairpin 'H'. **d**, The sequences of the transposon end, flanking donor-site DNA, and the cleaved transposon end and hairpin on the released donor-site DNA are shown.

separates the top strand of the flanking donor-site DNA from the top strand of the transposon. The same *Hermes*-L fragment is seen when the substrate 192-bp duplex fragment is isolated from a native gel and then run on a denaturing gel, indicating that the donor cleavage reaction begins with a nick near the 5' end of the transposon that precedes the double-strand break step (data not shown). The second product of the 'flank + *Hermes*-L' cleavage reaction migrates not at 101 nt, as would be expected from a simple double-strand break, but instead migrates more slowly than even the 194-nt substrate strands. Its apparent mobility of >200 nt is consistent with its being a DNA hairpin containing both strands of the flanking donor-site DNA (101 + 102 nt). When this slow-migrating species was isolated from a denaturing gel and rerun on a native gel, its mobility was at a position indicating a fragment of about 100 bp, consistent with a hairpin structure (data not shown).

To determine the structure of the hairpin species, we performed Maxam–Gilbert G sequencing (Fig. 2c). The G sequence of the hairpin species was identical to that of the bottom strand of the substrate flanking DNA with respect to positions G1, G6 and G10. However, the next G was not at position G16, as in the substrate, but rather at position G18, followed by Gs at G22, G25 and G30, reflecting the top strand sequence of the flanking donor-site DNA (Fig. 2d). This pattern is consistent with a hairpin that includes the bottom strand of the flanking donor-site DNA through positions T14 and A15 joined to A16 of the top strand of flanking DNA, and including the rest of the top strand of the flanking donor-site DNA. Analysis using Maxam–Gilbert G + A sequencing was consistent with this hairpin structure (data not shown). Additional evidence for a hairpin is that the T in the hairpin, T14, was hyper-reactive to potassium permanganate treatment (data not shown).

We conclude that *Hermes* donor cleavage involves two distinct chemical steps. Donor cleavage initiates with a nick one nucleotide into the donor strand that flanks the 5' end of the transposon, leaving one donor nucleotide attached to the 5' end of the transposon. The 3' OH of the top-flanking donor strand generated by this nick then joins to the bottom strand at the junction of the flanking donor-site DNA and the 3' end of the transposon. Hairpin

formation releases the 3' OH transposon terminus; the 5' transposon terminus is attached to one nucleotide of flanking donor-site DNA. This same mechanism is used in variable(diversity)joining recombination to promote the double-strand break reactions critical to the combinatorial assembly of variable(diversity)joining gene segments to produce diverse immunoglobulin and T-cell receptor genes in V(D)J recombination^{12,13}. However, in V(D)J recombination the double-strand break releases flush signal ends (analogous to the transposon ends) lacking unpaired 5' bases and generates a hairpin coding end (analogous to the donor site).

Hermes 3' OH termini join to target DNA

Hermes-L ends liberated by the hairpin-mediated breakage of the donor-site DNA join to a target DNA in the presence of *Hermes* transposase. On a native gel, target joining results in two new species (Fig. 3a, b). The lower species results from the concerted joining of two *Hermes*-L segments to the target plasmid, generating an appropriately sized linear plasmid; the upper species represents the joining of one *Hermes*-L transposon segment to the target plasmid, generating a nicked plasmid. As seen with the flanking donor hairpin, the amount of product resulting from joining of *Hermes*-L to the target plasmid increases over time.

To determine the chemistry of end joining to the target DNA, we examined the ability of *Hermes* transposase to join a pre-cleaved, that is, with its 3' OH terminus already exposed, *Hermes*-L fragment to target plasmids of two different sizes (Fig. 3c). On a native gel, species reflecting the joining of one and two *Hermes*-L fragments to the targets are observed (Fig. 3c, lanes 2, 3). On a denaturing gel, only a single unit-length product is evident with each target, consistent with the joining of one *Hermes*-L segment to a single target strand (Fig. 3c, lanes 5, 6).

Detection of these single-stranded products in reactions using a *Hermes*-L oligonucleotide labelled at the 5' end of its bottom strand reveals the chemistry of transposon end joining to the target: the 3' OH terminus of *Hermes*-L joins covalently to the target DNA. Interestingly, the 3' OH dinucleotide at the position of strand joining, TG-OH, which is highly conserved among *hAT* superfamily

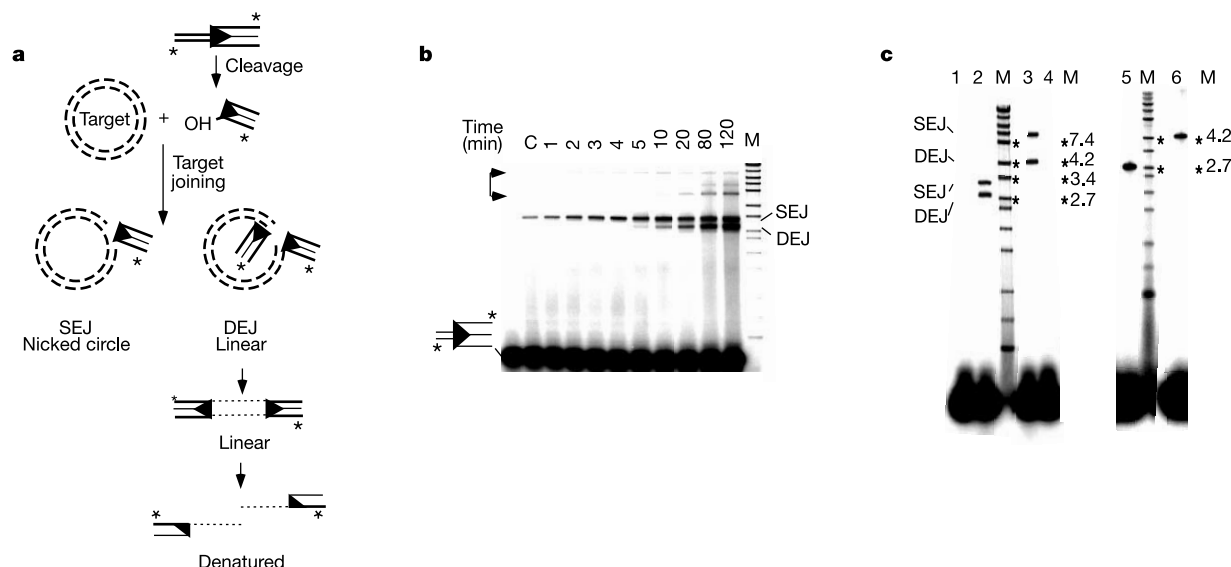


Figure 3 *Hermes* transposase joins the 3' OH end of *Hermes*-L to a target DNA. **a**, A schematic representation of the double-strand break that releases the transposon end from the flanking donor-site DNA and the subsequent joining of the transposon end to the target DNA is shown. SEJ, single-end join; DEJ, double-end join. **b**, The target-joining products generated over time when *Hermes* Tnsp is incubated in Mn^{2+} -containing buffer with a *Hermes*-L-containing fragment labelled at its 5' ends and a circular target plasmid (2.7 kb) as described in Fig. 2a are displayed on a native agarose gel. The 2.7-kb linear

product results from the joining of two *Hermes*-L fragments to the target plasmid; the 3.4-kb nicked circular product results from the joining of one *Hermes*-L to the target plasmid. Products with sizes >3.4 kb result from the joining of *Hermes* ends to oligomeric forms of the target plasmid. **c**, The target joining products generated when *Hermes* Tnsp is incubated with pre-cleaved *Hermes*-L ends radiolabelled at the 5' end of the bottom (transferred) strand and circular target plasmids of 2.7 and 4.2 kb are displayed on native (lanes 2, 3) and denaturing (lanes 5, 6) agarose gels. Lanes 1, 4: buffer control.

transposons⁶, including *Hermes*, is also found at the position of DNA cleavage and joining in V(D)J recombination^{12,14}.

Transposition by *Hermes* is accurate

We asked whether *Hermes* transposase could accurately join a DNA segment bounded by the *Hermes* termini to a target DNA. We generated a mini-*Hermes* element with pre-cleaved *Hermes*-L ends flanking a kanamycin resistance gene and used this DNA as a substrate for *in vitro* transposition into a target plasmid. Transposition products were recovered by transformation of *E. coli* with selection for kanamycin resistance. Junctions between the plasmid and the newly inserted transposons were sequenced. All 43 insertions analysed were accompanied by 8-bp target site duplication (data not shown) as occurs with *Hermes* and other *hAT* transposons *in vivo*¹⁵. Analysis of a mini-*Hermes* element containing *Hermes*-L and -R ends gave similar results (data not shown). Thus, our *Hermes* *in vitro* transposition system reflects *hAT* transposition *in vivo*.

Hermes transposition compared to other elements

Our experiments have revealed the mechanism of breakage and joining events underlying the translocation of a *hAT* superfamily transposon: the transposon is excised from the donor site by double-strand breaks mediated by hairpin formation on the flanking donor-site DNA, and the 3' OH transposon ends exposed by this cleavage join to the target DNA (Fig. 4). Element excision via double-strand breaks involving hairpin formation on the donor-site DNA has not been observed with any other transposon family. We note that although excision of Tn5 and Tn10 (members of the retroviral integrase superfamily) from their donor sites also involves hairpin formation^{16,17}, the hairpins in these systems are on the transposon ends, rather than on the flanking DNA as in *Hermes* transposition and V(D)J recombination. The strategy of joining of a 3' OH end to a target DNA is, however, highly conserved among mobile elements and can occur with the V(D)J recombinase RAG^{18,19} and members of the retroviral integrase superfamily³. It will be interesting to determine whether *Hermes*, like members of the retroviral integrase superfamily and RAG recombinases, uses direct one-step transesterification mechanisms both to promote hairpin formation^{20,21} and to join the transposon end to the target DNA^{21,22}. The experiments described below reveal that the same

active site in *Hermes* can mediate the distinct chemical steps of nicking, hairpin formation and target joining, as has been found with other transposases and integrases^{3,5}.

It will be especially interesting to determine the organization of the transposase complex that executes recombination: do single or multiple transposase protomers interact with each transposon end and, if there are multiple protomers, how are the chemical steps of recombination distributed among them? *Hermes* transposase can multimerize²³ (A.B.H. and F.D., unpublished work). Studies of Mu^{24,25} and Tn10²¹ transposition indicate that all the chemical steps in the movement of these elements are executed by a single transposase monomer at each transposon end, but the different order of the chemical steps and spatial positions of DNA substrates in *Hermes* and V(D)J recombination could well require multiple transposase active sites at each terminus.

After transposon excision, the broken donor chromosome must be repaired. With *hAT* elements, this occurs by rejoining of the exposed flanking DNA ends. The patterns of these joints led to the suggestion that hairpin intermediates were involved in *hAT* transposition^{26,27}; we have now verified that prediction biochemically. For repair to occur, these hairpins must be opened. Our experiments suggest that *Hermes* transposase does not perform this step, because the amount of hairpin formed increases over time (Fig. 2b) and no cleaved hairpin products have been detected. Therefore a host function like Artemis, which cleaves hairpins, is likely to be involved in *Hermes* transposition²⁸; an open reading frame (ORF) encoding an Artemis-like protein is present in the *Drosophila* and *Anopheles* genomes (Blast data not shown). Subsequent joining of the opened hairpins is probably performed by the host-encoded non-homologous end-joining DNA repair machinery^{27,29}. A potential active role for *Hermes* transposase in repair may be to maintain pairing of the broken DNAs to facilitate joining, a role suggested for the RAG recombinase in V(D)J recombination^{12,30,31}.

Unanticipated structure of the *Hermes* active site

Sequence comparisons among *hAT* elements have focused on genome-wide inspections and have identified amino-acid motifs that are often, but not always, present in *hAT* family members^{6,8}. To identify amino acids essential to *hAT* transposase activity, we

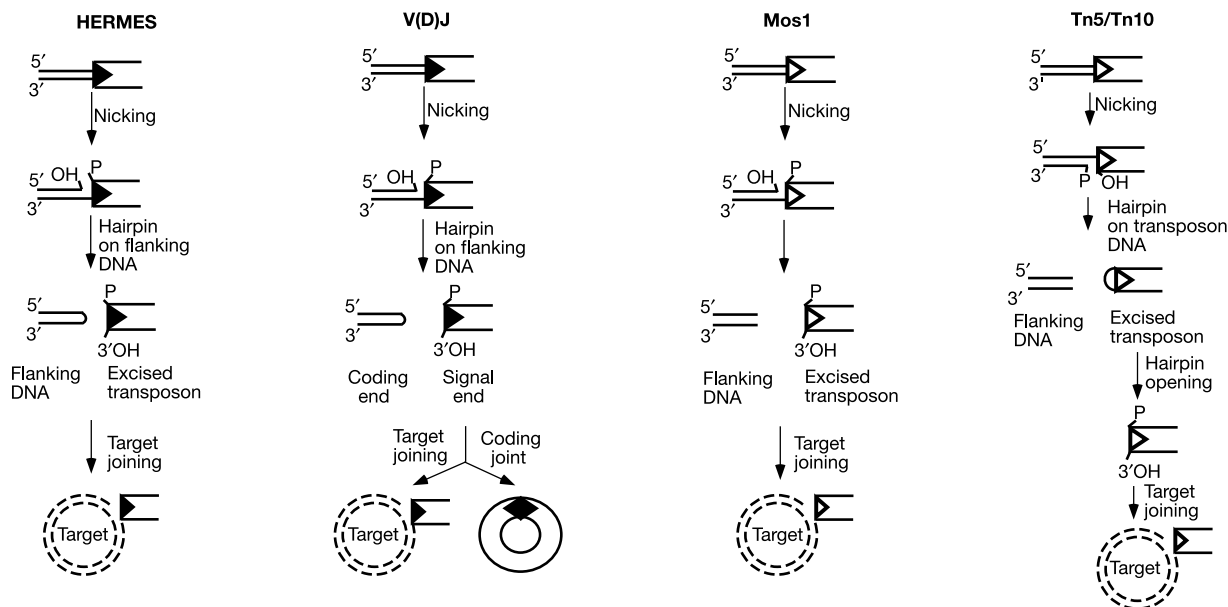


Figure 4 Comparison of recombinase-mediated cleavage and target-joining mechanisms. The positions of DNA cleavage reactions on various mobile DNAs are

shown. Although the exact positions of the 5' end cleavage vary, all elements are shown here with flush ends.

aligned the known active members of the *hAT* family using Clustal W³² (Supplementary Fig. 2) and identified several invariant acidic amino acids, in *Hermes* D180, D248 and E572, that could provide carboxylates to interact with the metal ion. The acidic DDE motif has previously been shown to be involved in the catalytic activity of members of the retroviral integrase superfamily and RAG recombinase^{3,12}. To probe the function of these amino acids in *Hermes* transposase, we purified mutant proteins with alanine substitutions at these positions and assayed them for DNA binding and recombination using several DNA substrates: (1) a substrate in which the transposon end was flanked by donor-site DNA, to probe coupled donor-site cleavage and strand transfer; (2) a 'pre-nicked' substrate with a nick in the top strand of the flanking donor-site DNA adjacent to the 5' end of the transposon, as is generated at the first step of *Hermes* transposition, to probe hairpin formation; and (3) a 'pre-cleaved' substrate in which the *Hermes* 3' OH ends were already exposed, to probe strand transfer.

Although still capable of binding specifically to *Hermes*-L (Supplementary Fig. 3), the D180A, D248A and E572A transposase mutants were inactive or had greatly reduced activity in almost all DNA breakage and joining steps (Fig. 5). These conserved acidic amino acids therefore play critical roles in catalysis. The fact that a single mutation can block multiple catalytic steps also suggests that a single active site mediates all the chemical steps of recombination. The mutants were greatly impaired at the first step in recombination, that is, single-strand nicking at the 5' end of the transposon; unlike in the wild type, no cleaved transposon top strand was seen after incubation of the mutants with a DNA substrate in which donor-site DNA flanked the transposon end with either Mn²⁺ or Mg²⁺ (Fig. 5a). Even when supplied with a 'pre-nicked' substrate, no DNA hairpin formation was observed with any mutant in Mg²⁺, although D248A was able to promote some hairpin formation in presence of Mn²⁺ (Fig. 5b). The alanine substitution mutants were also defective in strand transfer, because no joining of a 'pre-cleaved' transposon to a target DNA was detectable in Mg²⁺; some strand transfer with D248A was observed in the presence of Mn²⁺ (Fig. 5c).

We also examined the effects of cysteine substitutions at these DDE positions: specific suppression of a Cys mutant by Mn²⁺ has been associated with a direct interaction between the metal and amino acid^{33–35}. The presence of Mn²⁺ dramatically suppressed the

D180C defect in DNA hairpin formation and strand transfer, suggesting that this amino acid does interact with metal ion in the active site (Fig. 5b, c). We also observed some modest suppression by Mn²⁺ of the D248C defect in hairpin formation and target joining (Fig. 5b, c). The failure to observe Mn²⁺ suppression of E572C in any step of recombination or of D180C and D248C in a coupled cleavage and strand transfer assay (Fig. 5; Supplementary Fig. 4) probably reflects the fact that each metal binding site has a slightly different role in recombination³⁶.

Structural analysis of the HIV integrase and Mu and Tn5 transposases has previously demonstrated that the highly conserved catalytic DDE acidic amino acids lie on a common structure, that is, an RNaseH-like fold, and that they are closely juxtaposed, as appropriate for formation of a single active site^{5,37}. We performed secondary structure predictions with the *hAT* transposases from *Hermes*, *hobo*, *Ac* and *Tam3*³⁸. The results suggest that the conserved *hAT* catalytic acidic amino acids are also arranged in an RNaseH-like fold, suggesting that the active sites of *hAT* elements and retroviral integrase elements are actually similar (Supplementary Fig. 5). This finding is unexpected because no sequence homology between *hAT* transposases and retroviral integrase transposases has been detected⁶. The RAG1 subunit of the V(D)J recombinase also contains catalytic acidic amino acids that may also be organized in an RNaseH-like fold^{39–41}. Interestingly, with both *hAT* transposases and RAG1, and in contrast to other already known members of the retroviral integrase superfamily³, the second D and the E are widely separated in the sequence, D248–E572 in *Hermes* and D708–E962 in RAG1. It will be interesting to determine whether this wider spacing is functionally related to the fact that these enzymes make hairpins on the flanking DNA rather than on the transposon end as occurs with retroviral integrase elements Tn5 and Tn10.

Our identification of a retroviral-integrase-superfamily-like active site in what had been thought to be a distinct superfamily⁶ extends and broadens the types of mobile DNA elements that are likely to use the same mechanisms of DNA breakage and joining to transpose. Two other transposon superfamilies are *MuDR/Mu* of plants⁴² and the more widespread *piggyBac*⁴³. We have also analysed the secondary structures of members of these superfamilies (data not shown). The positions of conserved acidic amino acids suggests that at least part of the active sites of these transposases is related to

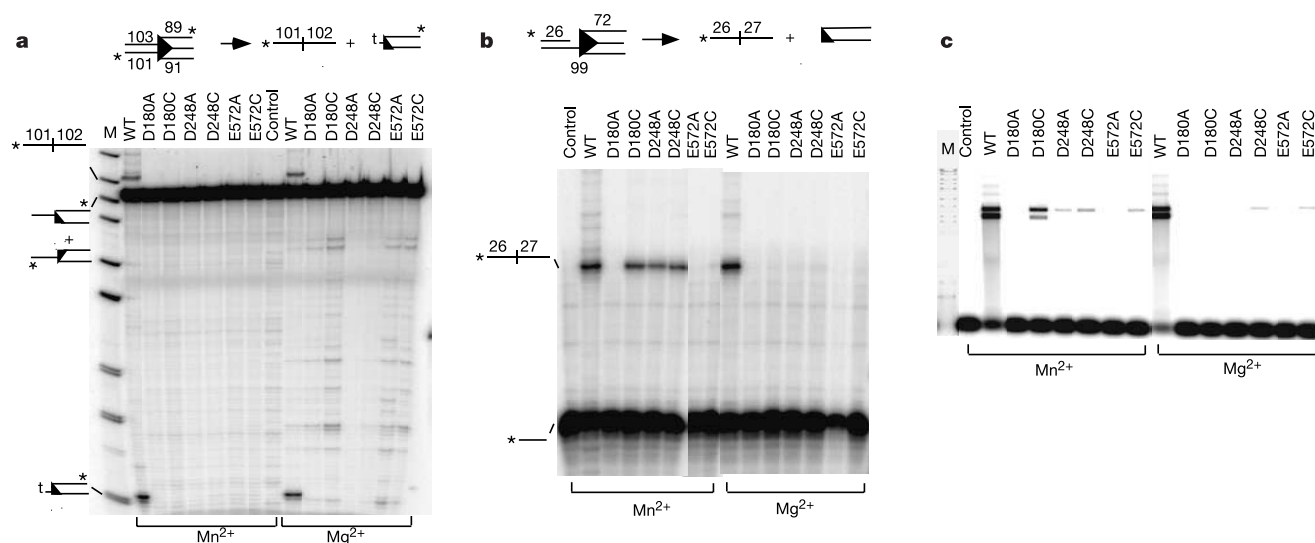


Figure 5 Mutation of *Hermes* DDE residues block DNA cleavage and strand transfer. **a, b**, The products of incubation of wild-type, alanine and cysteine substitutions at the DDE (conserved Asp-Asp-Glu) amino-acid motif of *Hermes* Tnsp in the presence of either Mn²⁺ or Mg²⁺ with various *Hermes*-L substrates are displayed along with schematic representations of each reaction. **a**, DNA nicking and hairpin formation using a

flank + *Hermes*-L substrate is visualized on a denaturing agarose gel. **b**, DNA hairpin formation using a pre-nicked flank + *Hermes*-L substrate is visualized on a denaturing acrylamide gel. **c**, DNA target joining using a pre-cleaved *Hermes*-L substrate is visualized on a native agarose gel.

those of the RNaseH-fold of the retroviral integrase and *hAT* superfamilies. Thus, very many superfamilies of elements all appear to contain similar active sites for DNA cleavage and strand transfer.

Other shared regions of *Hermes* and RAG1

In addition to the conserved DDE amino acids, there are several other interesting similarities between *Hermes* and RAG1. The formation of a DNA hairpin involves considerable DNA distortion. Such distortions can be facilitated by 'base flipping' during which a particular base becomes extrahelical, allowing distortion of the backbone⁴⁴. Base flipping can be stabilized by interactions between the base and aromatic amino acids (for example, tryptophan). A highly conserved region of *hAT* transposases⁶ contains a tryptophan, which is W319 in *Hermes* (Supplementary Fig. 2). Strikingly, there is a short region of amino-acid similarity between this region of *Hermes* sequence and the region of Tn5 transposase that contains a tryptophan involved in the base-flipping event that is essential to formation of the hairpin during Tn5 transposition^{45,46} (Supplementary Fig. 5). A highly conserved basic residue, R318 of *Hermes*, is closely juxtaposed to this tryptophan. We speculate that W319 of *Hermes* and related sequences in other *hAT* transposases will play key parts in DNA hairpin formation and base flipping. Mutational analysis of RAG1 has identified a number of basic amino acids whose alteration affects the hairpin formation step of RAG recombination⁴⁷; two of these, K889 and R894, flank a conserved tryptophan W893, and we speculate that W893 will also play a key role in hairpin formation, probably by promoting base flipping (Fig. 5). A similar role for this tryptophan has also been suggested on the basis of work on telomere resolvases⁴⁸.

Another very conserved feature of *hAT* transposases⁶ (Supplementary Fig. 2) is a CxxH motif (C265 and H268 in *Hermes*) downstream of the second catalytic D residue (D248 in *Hermes*); see Supplementary Figs 2 and 5. Mutation in H268 resulted in a severe defect in transposase catalytic activity (data not shown). RAG1 contains a C₂H₂ zinc-finger motif beginning at C720 just downstream of the second D, D708 (ref. 49).

A significant difference between *Hermes* transposition and RAG recombination is the fate of the DNA bound by the recognition and sites of action of the recombinases, that is, the TIRs of *Hermes* and the recombination signal sequences of V(D)J recombination. Recognition of these sites results in the DNA breakage that leaves hairpins of the flanking DNA and excises the segment bound by the TIRs or the recombination signal sequences. Whereas *Hermes* transposase often promotes insertion of the TIR-bounded fragment (the transposon) into a new target site after excision, the recombination-signal-sequence ends of a DNA fragment that excises from between the immunoglobulin and T-cell receptor genes to juxtapose novel gene coding segments almost always join to each other to form a 'signal sequence joint' and hence a circular species that is simply lost from the cell¹². Elucidation of how *Hermes* and RAG1 promote these different outcomes will lead to greater understanding of their mechanisms of action.

Conclusions

We have directly established the pathway of DNA breakage and joining reactions that underlie movement of a *hAT* family transposon and have also begun to dissect the structure and function of a *hAT* transposase. Our finding of an RNaseH-like fold with conserved acidic amino acids is surprising because the *hAT* family has long been considered to be quite distinct from the retroviral integrase superfamily. It had been suggested that the V(D)J recombination system may have evolved from an ancient transposable DNA element^{12,50}. Our findings here of such a close mechanistic relationship between *hAT* transposition and V(D)J recombination—that is, a double-strand break via hairpin formation on flanking DNA and 3' OH joining to the target DNA—and the related active sites of *hAT* transposases and RAG1 provides strong

support for the view that V(D)J recombination evolved from transposable elements. Dissection of the structure and function of *hAT* transposases will probably continue to provide important contributions to understanding V(D)J recombination. □

Methods

Hermes transposase expression and purification

The *Hermes* transposase (Tnsp) ORF (613 amino acids) was amplified by polymerase chain reaction (PCR) from plasmid pBCHSHH1.9 (ref. 11) and cloned between the *Nco*I and *Pvu*II sites of plasmid pBAD-Myc-HisB (Invitrogen) to generate a *Hermes*-Myc-His fusion construct. *Escherichia coli* strain Top10 (Invitrogen) transformed with the *Hermes*-Myc-His plasmid was grown overnight with shaking at 30 °C in LB medium containing 100 mg ml⁻¹ carbenicillin. The following day the overnight culture was diluted 1:100 with fresh LB+carbenicillin and cells were then grown to an absorbance at 600 nm of 0.6 at 30 °C. The culture was then shifted to 16 °C and induced with 0.1% L-arabinose for 16 h. After induction, cells were washed by centrifugation at 4 °C with TSG (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 10% v/v glycerol), and frozen in liquid nitrogen; all subsequent steps were performed at 4 °C. Frozen cells were resuspended in 10 ml TSG and lysed by sonication. The cleared lysate was loaded onto a pre-equilibrated Ni²⁺ Sepharose column (Amersham) and washed with ten column volumes of TSG, six column volumes of TSG+50 mM imidazole and six column volumes of TSG+100 mM imidazole. *Hermes*-Myc-His fusion protein was eluted with six column volumes of TSG+200 mM imidazole, dialysed against TSG, and stored at -80 °C.

DNA binding assay

DNA binding reactions were performed using *Hermes* end fragments generated by PCR from pBS*Hermes*^{10,11}. The 131-bp *Hermes*-L end fragment contained 81 bp *Hermes*-L+50 bp flanking housefly genomic DNA; the 178-bp *Hermes*-R fragment contained 110 bp *Hermes*-R+68 bp flanking housefly genomic DNA. DNAs were 5' end radiolabelled on both strands with γ -P³²-ATP. 140 nM *Hermes* Tnsp and 1.5 nM radiolabelled DNA were incubated in 25 mM HEPES, 3.5% (v/v) glycerol, 0.01% bovine serum albumin (BSA) and 4 mM dithiothreitol (DTT) in the presence of a 500-fold excess of sheared herring sperm DNA at room temperature for 20 min. Unlabelled 131-bp *Hermes*-L and 178-bp *Hermes*-R DNA was added in 50, 100 and 200 molar excess as indicated. The products were run on a 5% Tris-boric acid acrylamide gel at 4 °C. All gels were dried and exposed to phosphorimager plates and analysed by Imagequant software (Molecular Dynamics).

Analysis of coupled cleavage and strand transfer

Cleavage and strand transfer reactions were performed with a 193-bp *Hermes*-L end fragment containing 91 bp of *Hermes*-L+103-bp flanking housefly genomic DNA. The 194-bp *Hermes*-L fragment was radiolabelled on the 3' end of both strands with α -P³²-dATP. 140-nM *Hermes* transposase was incubated with 1.5 nM radiolabelled *Hermes*-L DNA in 25 mM HEPES, pH 7.6, 2% (v/v) glycerol, 50 mM NaCl, 2 mM DTT, 1 mM MnCl₂, 100 μ g ml⁻¹ BSA and 9.55 nM pUC19 plasmid as target DNA in a final volume of 20 μ l at 30 °C for the indicated times. Reactions were stopped by adding SDS and EDTA to 1% SDS and 20 mM EDTA and incubated for 1 h at 37 °C. The products were displayed on a 5% native acrylamide and on 1% agarose gels to look for double-strand break and target-joining products, respectively. For analysis of DNA nicking and hairpin formation, the reaction products were extracted with phenol/chloroform, precipitated with ethanol and run on a 5% urea acrylamide gel.

Analysis of hairpin formation using a pre-nicked *Hermes*-L end

A 26-bp oligonucleotide corresponding to the top-flanking donor-site DNA was radiolabelled at its 5' end with γ -P³²-ATP. After purification on a Sephadex G10 column, the end-labelled oligonucleotide was mixed with equimolar amounts of a 72-nt oligonucleotide corresponding to top strand of *Hermes*-L, and a 99-nt oligonucleotide containing 27 nt of bottom-strand flanking DNA+72 nt of bottom strand of *Hermes*-L (see Fig. 5b); the mixture was allowed to anneal and then used as substrate in coupled cleavage and strand-transfer reactions.

Strand-transfer reaction with pre-cleaved *Hermes*-L ends

Pre-cleaved *Hermes*-L for strand-transfer reactions was made by annealing oligonucleotides 5'-CAGAGAACAACAACAAGTGGCTTATTTGAGCTTGTCTGCG-3' (top) and 5'-CGCATAAGTATCAAAATAAGCCACTTGTGTGTTCTCTG-3' (bottom) radiolabelled at its 5' end with γ -P³²-dATP and used directly as a substrate for strand-transfer reactions with pUC19 and pBR322 target DNAs. Reactions were incubated for 2 h at 30 °C. The reactions were stopped by addition of SDS and EDTA to 1% SDS and 20 mM EDTA, incubated at 37 °C for 1 h, DNA was extracted with phenol/chloroform, precipitated with ethanol and loaded onto 1% TAE agarose and 1% NaOH agarose gels.

Recovery of transposon insertions

Primers NLC 1267 5'-CAGAGAACAACAACAAGTGGCTTATTTGAGCTTGTCTGCG-3' and NLC 1268 5'-CAGAGAACAACAACAAGTGGCTTATTTGACACTGGATGGCGGCG-3' were used to generate a kanamycin fragment flanked by 30 bp of *Hermes*-L at each end. These kanamycin fragments were used as substrate for a target-joining reaction with pUC19 as target DNA in the presence of *Hermes* Tnsp. The target-joining products were transformed into *E. coli*, kanamycin resistant colonies were selected, and the insertion site was sequenced.

Received 15 August; accepted 3 November 2004; doi:10.1038/nature03157.

- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53**, 245–281 (1999).
- Craig, N., Craigie, R., Gellert, M. & Lambowitz, A. *Mobile DNA II* (ASM, Washington DC, 2002).
- Rice, P. A. & Baker, T. A. Comparative architecture of transposase and integrase complexes. *Nature Struct. Biol.* **8**, 302–307 (2001).
- Rubin, E., Lithwick, G. & Levy, A. A. Structure and evolution of the *hAT* transposon superfamily. *Genetics* **158**, 949–957 (2001).
- Kunze, R. W. & Weil, C. F. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A.) 565–610 (ASM, Washington DC, 2002).
- Robertson, H. M. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A.) 1093–1110 (ASM, Washington DC, 2002).
- Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. Evidence for a common evolutionary origin of inverted repeat transposon in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell* **66**, 465–471 (1991).
- Warren, W. D., Atkinson, P. W. & O'Brochta, D. A. The Hermes transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genet. Res.* **64**, 87–97 (1994).
- O'Brochta, D. A., Warren, W. D., Saville, K. J. & Atkinson, P. W. Hermes, a functional non-Drosophilid insect gene vector from *Musca domestica*. *Genetics* **142**, 907–914 (1996).
- Gellert, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101–132 (2002).
- Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J. & Gellert, M. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in SCID mouse thymocytes. *Cell* **70**, 983–991 (1992).
- Okuda, M., Ikeda, K., Namiki, F., Nishi, K. & Tsuge, T. Tfo1: an *Ac*-like transposon from the plant pathogenic fungus *Fusarium oxysporum*. *Mol. Gen. Genet.* **258**, 599–607 (1998).
- Sarkar, A. *et al.* The Hermes element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetics* **99**, 15–29 (1997).
- Kennedy, A., Guhathakurta, A., Kleckner, N. & Haniford, D. B. Tn10 transposition via a DNA hairpin intermediate. *Cell* **95**, 125–134 (1998).
- Bhasin, A., Goryshin, I. Y. & Reznikoff, W. S. Hairpin formation in Tn5 transposition. *J. Biol. Chem.* **274**, 37021–37029 (1999).
- Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
- Agrawal, A., Eastman, O. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
- van Gent, D. C., Mizuuchi, K. & Gellert, M. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592–1594 (1996).
- Kennedy, A. K., Haniford, D. B. & Mizuuchi, K. Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell* **101**, 295–305 (2000).
- Mizuuchi, K. & Adzuma, K. Inversion of the phosphate chirality at the target site of the Mu DNA strand transfer: evidence for a one-step transesterification mechanism. *Cell* **66**, 129–140 (1991).
- Michel, K., O'Brochta, D. A. & Atkinson, P. W. The C-terminus of the Hermes transposase contains a protein multimerization domain. *Insect Biochem. Mol. Biol.* **33**, 959–970 (2003).
- Namgoong, S. & Harshey, R. The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. *EMBO J.* **17**, 3775–3785 (1998).
- Williams, T. L., Jackson, E. L., Carritte, A. & Baker, T. A. Organization and dynamics of the Mu transpososome: recombination by communication between two active sites. *Genes Dev.* **13**, 2725–2737 (1999).
- Coen, E. S., Carpenter, R. & Martin, C. Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47**, 285–296 (1986).
- Weil, C. F. & Kunze, R. Transposition of maize *Ac/Ds* transposable elements in the yeast *Saccharomyces cerevisiae*. *Nature Genet.* **26**, 187–190 (2000).
- Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).

- Yu, J., Marshall, K., Yamaguchi, M., Haber, J. E. & Weil, C. F. Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **24**, 1351–1364 (2004).
- Qiu, J. X., Kale, S. B., Yarnell, S. H. & Roth, D. B. Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination. *Mol. Cell* **7**, 77–87 (2001).
- Lee, G., Neiditch, M., Salus, S. & Roth, D. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* **117**, 171–184 (2004).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Sarnovsky, R., May, E. W. & Craig, N. L. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996).
- Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* **361**, 85–88 (1993).
- Allingham, J., Pribil, P. & Haniford, D. All three residues of the Tn10 transposase DDE catalytic triad function in divalent metal ion binding. *J. Mol. Biol.* **289**, 1195–1206 (1999).
- Brautigam, C. & Steitz, T. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54–63 (1998).
- Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
- McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
- Kim, D. R., Dai, Y., Mundy, C. L., Yang, W. & Oettinger, M. A. Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev.* **13**, 3070–3080 (1999).
- Fugmann, S. D., Villey, I. J., Ptazek, L. M. & Schatz, D. G. Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol. Cell* **5**, 97–107 (2000).
- Landree, M. A., Wibbenmeyer, J. A. & Roth, D. B. Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev.* **13**, 3059–3069 (1999).
- Walbot, V. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A.) 533–563 (ASM, Washington DC, 2002).
- Sarkar, A. *et al.* Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. *Mol. Genet. Genom.* **270**, 173–180 (2003).
- Roberts, R. & Cheng, X. Base flipping. *Annu. Rev. Biochem.* **67**, 181–198 (1998).
- Davies, D. R., Braam, L. M., Reznikoff, W. S. & Rayment, I. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9 Å resolution. *J. Biol. Chem.* **274**, 11904–11913 (1999).
- Ason, B. & Reznikoff, W. S. Mutational analysis of the base flipping event found in Tn5 transposition. *J. Biol. Chem.* **277**, 11284–11291 (2002).
- Huye, L. E., Purugganan, M. M., Jiang, M. M. & Roth, D. B. Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. *Mol. Cell. Biol.* **22**, 3460–3473 (2002).
- Bankhead, T. & Chaconas, G. Mixing active-site components: a recipe for the unique enzymatic activity of a telomere resolvase. *Proc. Natl Acad. Sci. USA* **101**, 13768–13773 (2004).
- De, P. & Rodgers, K. Putting the pieces together: identification and characterization of structural domains in the V(D)J recombination protein RAG1. *Immunol. Rev.* **200**, 70–82 (2004).
- Lewis, S. M. & Wu, G. E. The origins of V(D)J recombination. *Cell* **88**, 159–162 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to G. Chaconas for sharing his thoughts about hairpin binding pockets before publication. N.L.C. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.L.C. (ncraig@jhmi.edu).

Cavity cooling of a microlever

Constanze H hberger Metzger & Khaled Karrai

Center for NanoScience and Sektion Physik, Ludwig-Maximilians-Universit t, Geschwister-Scholl-Platz 1, 80539 M nchen, Germany

The prospect of realizing entangled quantum states between macroscopic objects and photons¹ has recently stimulated interest in new laser-cooling schemes^{2,3}. For example, laser-cooling of the vibrational modes of a mirror can be achieved by subjecting it to a radiation² or photothermal⁴ pressure, actively controlled through a servo loop adjusted to oppose its brownian thermal motion within a preset frequency window. In contrast, atoms can be laser-cooled passively without such active feedback, because their random motion is intrinsically damped through their interaction with radiation^{5–8}. Here we report direct experimental evidence for passive (or intrinsic) optical cooling of a micro-mechanical resonator. We exploit cavity-induced photothermal pressure to quench the brownian vibrational fluctuations of a gold-coated silicon microlever from room temperature down to an effective temperature of 18 K. Extending this method to optical-cavity-induced radiation pressure might enable the quantum limit to be attained, opening the way for experimental investigations of macroscopic quantum superposition states¹ involving numbers of atoms of the order of 10¹⁴.

The gold-coated silicon micro-cantilever forms one of the two mirrors of a miniature Fabry–P rot (FP) optical cavity in a set-up shown in Fig. 1. When the FP cavity is tuned near one of its optical resonances, the density of photons stored in it is resonantly enhanced. Equilibrium is reached only after a time when the flux of stored photons equals the photon leakage, which is entirely determined by the mirrors' transmission as well as the mirror separation. As the lever is mechanically flexible, it moves under the effect of the resonantly increased photon-induced forces. So when subjected to a strong enough laser field, the lever mirror displacement affects in turn the photon density stored in the FP cavity. In this way the lever mechanics is strongly coupled to the light field. Such a concept has been explored in pioneering work^{9,10} and later verified with laser fields¹¹. We recently demonstrated that the effect is also seen in lever-based microcavities¹². During the lever motion, the photon leakage rate in the FP cavity changes as well. Because this change is not instantaneous but delayed, it leads to a force experienced by the lever that is proportional to its velocity⁹. Here we demonstrate experimentally that such a light-induced viscous damping is responsible for the cavity cooling of the lever.

First we show how laser-induced cooling is monitored. As the brownian motion of the lever is a measure of its temperature, we have analysed quantitatively the noise spectrum in the cantilever thermal-induced motion in the vicinity of the lowest lever vibrational resonance frequency at 7.3 kHz. As attested by the remarkable behaviour shown in Fig. 2a (obtained for a slightly blue detuned FP cavity using a lever placed in vacuum), we found that increasing laser power reduces the amplitude of the brownian motion by almost two orders of magnitude. In other words, the lever behaves as if it were refrigerated. The advantage of this method is that the lever geometry, the materials used and the FP cavity size are parameters that can be engineered to optimize cavity cooling.

We derive now a model that accounts quantitatively for the observed cooling of the vibrational mode. The displacement z of the cantilever from its equilibrium position obeys Newton's equation of motion. As the cantilever has an elastic rigidity K , its equation of motion for small amplitudes z can be approximated by that of a driven damped harmonic oscillator with an effective mass¹³ m . An effective thermal force F_{th} in the equation of motion¹⁴ drives

the lever into brownian motion. Crucial to the present cooling scheme is the fact that the lever is also subjected to photon-induced forces that depend on the FP cavity length and are hence a function of the lever displacement z . The photon-induced force $F(z) = \sum_n F_n(z)$, which is assumed proportional to the light intensity stored in the cavity, includes of course the radiation pressure but more generally all the n independent light-induced contributions, such as the photothermal (bolometric), radiometric and photo-elastic pressure to name just a few. For instance, a bolometric force F_B results from the differential thermal expansion between the silicon lever and the thin gold film (refs 12, 15 and 16, and ref. 17 and refs therein). When photons are absorbed in the gold film, the composite gold/silicon lever heats locally and bends, thereby changing the cavity length. As we will see below, the essence of cooling is based on the fact that the optically induced forces acting on the lever are delayed with respect to a sudden change in the lever position. To take this delayed response into account, we define $h_n(t - t')$, which is the response function that gives the force contribution F_n at time t for a sudden change $z(t')$ at time t' . The equation of motion is

$$m \frac{d^2 z}{dt^2} + m \Gamma \frac{dz}{dt} + Kz = F_{th} + \sum_n \int_0^t \frac{dF_n[z(t')]}{dt} h_n(t - t') dt' \quad (1)$$

where the damping factor Γ accounts for the mechanical energy loss inherent to the microlever. We have assumed exponential delayed responses, $h_n(t) = 1 - \exp(-t/\tau_n)$, where τ_n is the time it takes for the light-induced force F_n to reach equilibrium after a sudden change in z .

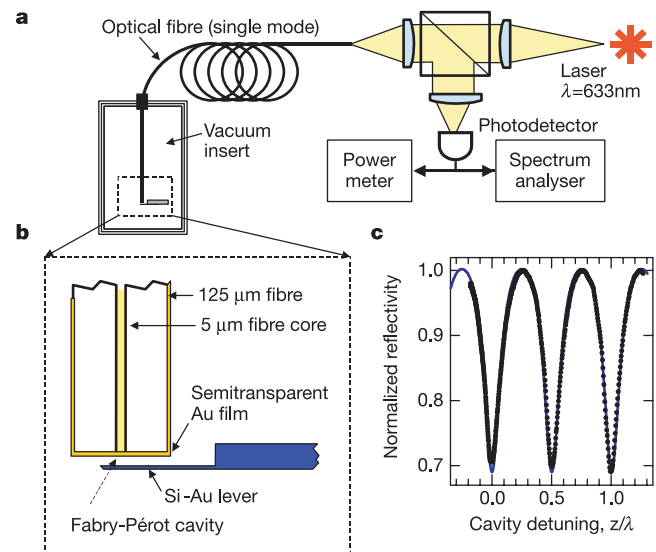


Figure 1 Experimental set-up. **a**, The light of a HeNe laser with wavelength $\lambda = 633$ nm is launched into a single-mode fibre. The other end of the fibre is in a vacuum chamber (10^{-6} mbar)—this end is polished and coated with a gold semitransparent thin film. **b**, The flat end of the fibre makes one of the two mirrors of a Fabry–P rot cavity. The second mirror is made out of a thin silicon microlever 223 μm long, 22 μm broad and 0.46 μm thick, corresponding to a nominal elastic constant of $K = 0.008 \text{ N m}^{-1}$. Both faces of the cantilever are coated with ~ 35 -nm-thick gold films, so that a slight asymmetry in the film thickness leads to a photo-thermal induced force. The nominal separation between the two mirrors is $\sim 34 \mu\text{m}$. The light reflected from the FP cavity retraces its way back through the fibre, through a beam splitter to a photodetector. In order to avoid spurious effects, the laser was optically isolated from the lever so that less than 1 in 10^5 parts of the reflected signal returned to the source. **c**, The measured reflected light power plotted as a function of the mirror distance shows FP resonances (dots). The line through the data is the best fit to a model for a cavity with both mirrors reflecting 40% each. The cavity detuning length z is adjusted electrostatically by applying a DC voltage between the gold film coating the fibre and the lever¹². A spectrum analyser (0–100 kHz) is used to acquire the thermal noise spectrum in the region close to the lowest vibrational resonance of the lever.

We solve equation (1) in the limit of small lever amplitudes z typical of brownian fluctuations. In this case, the photon-induced force can be approximated by its two first terms in Taylor expansion around the lever's equilibrium position, namely $F(z) \approx \Sigma_n (F_{0n} + [\partial F/\partial z]_{0n} z)$. The first term is the optical driving force, the DC component of which displaces the lever to a new equilibrium position. The second term, which can be positive or negative depending on the cavity detuning^{9,12}, is proportional to the lever displacement. This means that the second term can be identified with elastic forces with photon-induced rigidities^{9,12} $K_n = -[\partial F/\partial z]_{0n}$ that add to the lever mechanical elastic rigidity K . $K_n = 0$ for cavities exactly tuned on optical resonance, and K_n extremal values (which are proportional to the light intensity and the third power of the cavity finesse¹²), are obtained by slightly detuning the cavity. The equation of motion decomposed in the frequency domain f has then a simple solution for the lever amplitude components z_ω at cyclic frequency $\omega = 2\pi f$ given by

$$z_\omega = \frac{\sum_n \Delta F_{0n,\omega} / (1 + i\omega\tau_n) + F_{th,\omega}}{K} \frac{\omega_0^2}{\omega_{eff}^2 - \omega^2 + i\Gamma_{eff}\omega} \quad (2)$$

where $\Delta F_{0n,\omega}$ and $F_{th,\omega}$ are the Fourier components of a weakly modulated driving force ΔF_{0n} and of F_{th} , respectively. $\omega_0^2 = K/m$ is the lowest vibrational resonance of the lever in the absence of illumination. The above expression is that of the amplitude spectrum of a driven harmonic oscillator with an optically modified damping Γ_{eff} and a modified resonance frequency ω_{eff} given by

$$\Gamma_{eff} = \Gamma \left(1 - Q_M \sum_n \frac{\omega_0 \tau_n}{\omega^2 \tau_n^2 + 1} \frac{K_n}{K} \right) \quad (3)$$

$$\omega_{eff}^2 = \omega_0^2 \left(1 + \sum_n \frac{1}{\omega^2 \tau_n^2 + 1} \frac{K_n}{K} \right) \quad (4)$$

where $Q_M = \omega_0/\Gamma$ is the mechanical quality factor of the lever vibrational resonance, which here is typically of the order of 2×10^3 in vacuum, but as shown in ref. 18, it can be made as large as 2×10^5 .

From equation (3), we see that the photo-induced change in the viscous damping is controlled through the light-induced force rigidity K_n and the delay τ_n . In particular, when operating on cavity

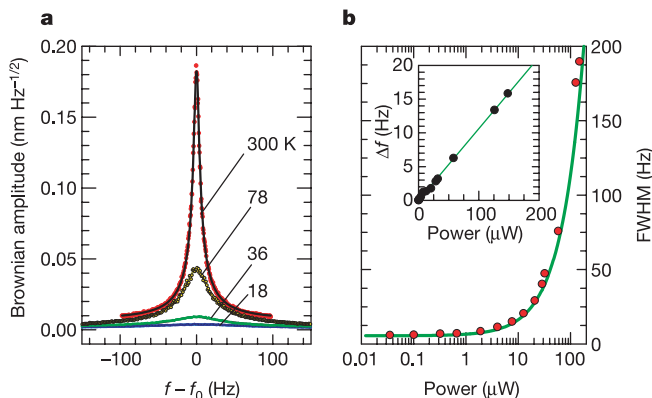


Figure 2 Cavity-induced cooling of the lever vibrational resonance. **a**, Brownian motion amplitude spectra of the lever vibrational resonance measured for four different laser power levels. The largest peak spectrum data (dots) centred on $\omega_0 = 2\pi \times 7,280$ Hz was obtained at 80 nW. The full line is the calculated brownian noise spectrum using a temperature $T = 300$ K. Increasing the laser power to 130 μ W reduced markedly the area of the peak and shifted its frequency to $\omega_{eff} = 2\pi \times 7,300$ kHz. The spectra calculated using equations (5) and (7) fit perfectly the data obtained at three different increasing laser powers when effective temperatures $T_{eff} = 78, 36$ and 18 K are assumed. The cantilever amplitude calibration is obtained independently using the measured interferogram shown in Fig. 1c. The FP cavity was $+\lambda/25$ detuned from one of its optical resonances. **b**, Full-width at half-maximum (FWHM) of the vibrational resonance as a function of the power reflected from the cavity. The full line is a fit according to equation (3). Inset: frequency shift of the resonance. The full line is a fit according to equation (4).

detuning such that $K_n < 0$, equation (3) shows that Γ_{eff} increases linearly with K_n and hence with the photon intensity in the cavity. This behaviour is directly experimentally confirmed in Fig. 2b, where the linewidth of the vibrational resonance is seen to increase with the laser power coupled into the cavity. A further consistency check is obtained by changing the sign of detuning such that $K_n > 0$. In this situation, a value of the product $Q_M \tau_n K_n$ can be reached that nulls the effective damping Γ_{eff} and even makes it negative. When such a situation is reached, the lever is no longer damped and thermal fluctuations are strong enough to set it into a mode of self-oscillation. Such instabilities were first explored in refs 9 and 10 and are often discussed in relation to large optomechanical systems, such as a gravitational wave antenna¹⁹. Under a condition of red detuning, we indeed observed (just as predicted) that above a threshold of laser intensity the lever enters into a mode of self-oscillation at the resonance frequency. The corresponding resonance is seen in Fig. 3c.

The lever was placed in high vacuum so that only the radiation and bolometric pressure are expected to dominate. In order to determine their relative contribution, we measured the full spectrum of z_ω (not shown) over five decades in frequency, from 1 Hz to 100 kHz, by weakly modulating the laser intensity. In our particular device, the distance between both mirrors (34 μ m) is short enough, and the mirrors' reflectivities are low enough, that a change in the lever position leads to an instantaneous change in the photon intensity in the cavity. Therefore, the radiation pressure changes quasi-instantaneously with fluctuations of the lever position, and the cavity-induced damping seen in Fig. 2b is dominated by the bolometric contribution alone. Fitting equation (2) to the measured

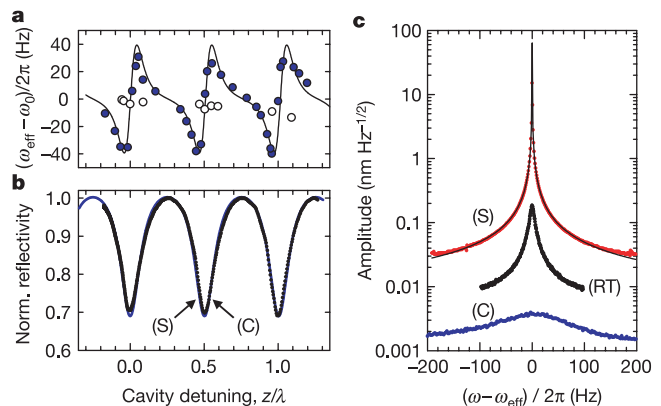


Figure 3 Lever vibrational resonance frequency and shape as a function of the optical cavity tuning. **a**, Frequency shift of the lowest mechanical lever resonance as a function of the FP cavity detuning (here measured at room pressure). The data (filled circles) measured at 100 μ W laser power show a clear periodic dependency with the cavity length, demonstrating optical control of the mechanical resonance. The FP reflectivity resonances shown for direct comparison in **b** have the same periodicity. In contrast, the data (open circles in **a**) measured at lower laser power (5 μ W) show no effect. The vibrational resonance frequencies are obtained by fitting the spectral noise amplitude to equation (5). The fit (full line in **a**) is given using equation (4) together with the photon-induced rigidity $K_{ph} = kF_0(g^3/\sqrt{R})\text{sinc}kz_0\text{cos}kz_0/(1 + g^2\text{sinc}kz_0)^2$ from ref. 12, where $k = 2\pi/\lambda$, F_0 is the light-induced force that the lever would experience in the absence of the FP cavity, z_0 is the equilibrium mirror separation, and $g = 2\sqrt{R}/T$ is the FP cavity finesse assuming that both mirrors have a transmission R and reflectivity T . **c**, Brownian amplitude noise spectrum of the lever near the lowest vibrational resonance frequency, measured under vacuum conditions with 130 μ W laser power. Cooling, demonstrated in curve (C), is obtained by detuning the cavity length by about $+\lambda/25$ from an FP optical resonance (working point shown by an arrow labelled (C) in **b**). Self-excitation, shown by curve (S), in contrast, is obtained by detuning the cavity length by about $-\lambda/25$ (working point (S) indicated by arrow in **b**). The curve labelled (RT) is the room-temperature brownian motion amplitude obtained at 80 nW laser power. The data in **c** are fitted using equation (5) together with equations (3) and (4).

z_ω spectrum, we obtained the response time of the bolometric force, $\tau_B = 560 \mu\text{s}$, and the ratio of the bolometric to radiation pressure contribution, $F_{B0}/F_{R0} = K_B/K_R = -95$, indicating that for this particular gold-coated lever, bolometric forces and radiation pressure act in opposite directions. Investigating the behaviour of the vibrational resonance frequency further strengthens the consistency of our analysis. Equation (4) shows that all light rigidities K_n contribute to modify the lever resonance frequency. However, in the present case for the frequency range centred on the lever vibrational resonance frequency $\omega \approx \omega_0 = 2\pi \times 7.3 \text{ kHz}$, the denominator $(\omega_0\tau_B)^2 + 1 \approx 626$ reduces the bolometric contribution so much that the light-induced shift in the effective resonance frequency ω_{eff} is fully dominated by the radiation pressure. Tuning the cavity length changes periodically the radiation-pressure-induced rigidity K_R from positive to negative values, increasing or decreasing the mechanical resonance frequency, just as observed in Fig. 3a.

Damping does not necessarily imply cooling. To see how the temperature can be lowered, we analyse the effects of cavity-induced damping on the brownian vibrational motion of the lever. When the laser source intensity is kept constant in time ($\Delta F_{0n,\omega} = 0$ for $\omega \neq 0$), the lever is only thermally driven and its mechanical resonance is revealed in the noise spectrum of the light reflected from the microcavity, as shown in Fig. 2. The effective lever temperature is obtained by considering the mean square of the spectral components of the thermal driving force¹⁴ F_{th} in a frequency window $\delta f = \delta\omega/2\pi$. The result gives the frequency component of the mean of the lever's squared amplitude:

$$\langle z_\omega^2 \rangle = \frac{4k_B T}{K} \frac{\omega_0^2 \Gamma}{(\omega_{\text{eff}}^2 - \omega^2)^2 + (\Gamma_{\text{eff}} \omega)^2} \delta f \quad (5)$$

Integrating it over the whole frequency spectrum, we obtain the mean of the lever noise squared amplitude, which relates directly to the lever mean kinetic energy. In our experimental situation, $\omega_{\text{eff}} \approx \omega_0$ and $\Gamma_{\text{eff}} \ll \omega_0$, the result approximates to

$$\frac{1}{2} K \langle z^2 \rangle \cong \frac{1}{2} k_B T \left(\frac{\omega_0^2}{\omega_{\text{eff},0}^2} \frac{\Gamma}{\Gamma_{\text{eff},0}} \right) \quad (6)$$

where $\omega_{\text{eff},0}$ and $\Gamma_{\text{eff},0}$ are the values of ω_{eff} and Γ_{eff} given in equations (3) and (4) setting $\omega = \omega_0$. This expression can be identified with the equipartition theorem for a one-dimensional harmonic oscillator at an effective temperature given by:

$$T_{\text{eff}} = T \frac{\omega_0^2}{\omega_{\text{eff},0}^2} \frac{\Gamma}{\Gamma_{\text{eff},0}} \quad (7)$$

This expression demonstrates that a control of the mechanical oscillator's temperature is obtained through cavity tuning of both the effective vibrational resonance frequency and the effective damping, and our measurements provide a direct evidence of its validity. Our analysis is generalized to any arbitrary delayed force, and it is not limited to the bolometric pressure alone. It can be shown that for any given type of photon-induced force $F_n(z)$, the condition for optimal damping is obtained when both the photon-induced rigidity equals the lever rigidity (that is, $K_n = -K$) and the time constant of the force is of the order of the lever period (that is, $\tau_n = \sqrt{2}/\omega_0$). Ideally then, effective temperature can be lowered to a minimum of $T_{\text{eff}}/T = 4\sqrt{2}/Q_M$. The best cooling is therefore obtained for large Q_M . In vacuum and at low temperatures, Q_M can be made as large as 10^5 , which means that sub-millikelvin effective temperatures are in principle possible when starting from a surrounding temperature in the sub-kelvin range.

In our particular example, for which cooling is controlled through bolometric effects, the brownian motion data measured in the laser-cooling regime in Fig. 2, and analysed with equation (5) and (7), correspond to a reduced temperature as low as $T_{\text{eff}} \approx 18 \text{ K}$

down from room temperature. In principle, the limit of cavity cooling is reached when the heating resulting from the light power absorbed in the lever mirror exceeds the cooling power. In our experiment, this situation was still not reached for an effective cooling of 18 K . Increasing the cavity reflectivity increases the contribution of radiation pressure to the optical damping. This is because in a low-loss FP cavity with a mirror separation L and with high reflectivity R (that is, high finesse), the radiation pressure does not change instantaneously with the lever motion—instead it is delayed by the cavity storage time $\tau_R = L/[c(1-R)]$. In such an ideal cavity, the lever cooling is analogous to that of atomic molasses. The analogy is revealed when we compare the expression of the velocity-dependent forces determined for (1) an atom placed in counter-propagating laser beams and (2) for the lever subject to the retarded radiation pressure in the cavity. In both cases, we determine that the optically induced damping rate at frequency ω is given by

$$\Gamma_{\text{opt}} = \frac{S_0 \sigma}{mc^2} \frac{\Omega \tau}{\omega^2 \tau^2 + 1} \left\{ \frac{4\delta\tau}{(1 + 4\tau^2 \delta^2)^2} \right\} \quad (8)$$

where S_0 is the laser total power density, $\Omega = 2\pi\nu$, where ν is the photon frequency with wavelength λ , and m is the atom or lever mass. The delay time τ corresponds to the radiative lifetime of the atom optical transition, whereas for the lever it is the cavity storage time τ_R . Here σ has the dimension of area. For atoms, it is the optical scattering cross-section at the optical transition frequency, namely $\sigma = 3\lambda^2/2\pi$. For the FP cavity, the cross-section is instead $\sigma \approx 8RA/(1-R)^2$, where A is the illuminated area on the lever.

The term in parentheses in equation (8) is simply a numerical function of δ , the frequency detuning from optical resonance. It takes the value 0 for $\delta = 0$ on cavity resonance, and has the maximum value $+3\sqrt{3}/8$ for a red-shift detuning $\delta_- = -1/(2\sqrt{3}\tau)$. At this detuning under a laser power P_0 , we derived that the lever vibrational effective temperature is:

$$T_{\text{eff}} = \frac{T}{\left[1 + 6\pi\sqrt{3} \frac{P_0/\Gamma}{mc^2} \frac{L R(1+R)}{\lambda(1-R)^3} \frac{1}{1+\omega_0^2 \tau_R^2} \right] \left[1 - \frac{3\sqrt{3} P_0/\omega_0}{2 mc^2} \frac{\Omega}{\omega_0(1-R)^2} \frac{R\sqrt{R}}{1+\omega_0^2 \tau_R^2} \right]} \quad (9)$$

In the denominator, the first term in brackets originates from the cavity-induced damping, while the second term is due to the cavity-induced rigidity. For $T_{\text{eff}} \ll T$, the optimal working point is given at frequency $\omega_0\tau_R \approx 1$, namely $\omega_0 \approx c(1-R)/L$, and the lever thermal energy in units of the vibrational quantum is then minimized down to:

$$\frac{k_B T_{\text{eff}}}{\hbar\omega_0} \cong \frac{2}{3\sqrt{3}} \frac{k_B T}{\hbar\Omega} \frac{mc^2}{P_0/\Gamma} \frac{(1-R)^2}{R} \quad (10)$$

The quantum limit is reached for $k_B T_{\text{eff}} < \hbar\omega_0$. Using a mechanical resonator (with a mass $m = 2.5 \times 10^{-15} \text{ kg}$ and a natural damping rate of $\Gamma \approx 20 \text{ Hz}$) made of a tiny $1 \times 1 \times 1 \mu\text{m}^3$ sized Bragg dielectric mirror with $R = 0.99$ placed in a cavity with $L = 20 \text{ cm}$ and illuminated with $P_0 = 2 \mu\text{W}$ and $\lambda = 1 \mu\text{m}$ would lead to a temperature reduction of $1/5,000$ at 400 kHz . For a lever pre-cooled to $T = 100 \text{ mK}$, the added optical cooling would leave it in its quantum limit should its vibrational frequency be between 400 kHz and 10 MHz . Here, as in the case of bolometric cooling, the potential limitation is reached when the residual optical absorption in the lever heats it more than it cools it. Whether the quantum limit or the residual optical thermal heating of the vibrational mode is reached first remains to be experimentally investigated. $\square$

Received 27 June; accepted 18 October 2004; doi:10.1038/nature03118.

1. Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. Toward quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401 (2003).
2. Cohadon, P. F., Heidmann, A. & Pinard, M. Cooling of a mirror by radiation pressure. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).
3. Wilson-Rae, I., Zoller, P. & Imamoglu, A. Laser cooling of a nanomechanical resonator mode to its

quantum ground state. *Phys. Rev. Lett.* **92**, 075507 (2004).

4. Mertz, J., Marti, O. & Mlynek, J. Regulation of a microlever responses by force feedback. *Appl. Phys. Lett.* **62**, 2344–2346 (1993).
5. Hänsch, T. W. & Schawlow, A. Cooling of gases by laser radiation. *Opt. Commun.* **13**, 68–69 (1975).
6. Cohen-Tannoudji, C. & Phillips, W. New mechanisms for laser cooling. *Phys. Today* **43**, 33–52 (1990).
7. Chu, S., Hollberg, L., Björholm, J. E., Cable, A. & Ashkin, A. Three-dimensional viscous confinement and cooling of atoms by resonance radiation pressure. *Phys. Rev. Lett.* **55**, 48–51 (1985).
8. Epstein, R. L., Buchwald, M. L., Edwards, B. C., Gosnell, T. R. & Mungan, C. E. Observation of laser-induced fluorescent cooling of a solid. *Nature* **377**, 500–502 (1995).
9. Braginsky, V. B. & Manukin, A. B. *Measurements of Weak Forces in Physics Experiments* (Chicago Univ. Press, Chicago, 1977).
10. Braginsky, V. B., Manukin, A. B. & Tikhonov, M. Yu. Investigation of dissipative ponderomotive effects of electromagnetic radiation. *Zh. Eksp. Teor. Fiz.* **58**, 1549–1552 (1970); *Sov. Phys. JETP* **31**, 829–830 (1970).
11. Dörsel, A., McCullen, J. D., Meystre, P., Vignes, E. & Walther, H. Optical bistability and mirror confinement induced by radiation pressure. *Phys. Rev. Lett.* **51**, 1550–1553 (1983).
12. Vogel, M., Mooser, C., Karrai, K. & Warburton, R. Optically tunable mechanics of microlevers. *Appl. Phys. Lett.* **83**, 1337–1339 (2003).
13. Sarid, D. *Scanning Force Microscopy* (Oxford Univ. Press, New York, 1991).
14. Gillespie, D. T. The mathematics of Brownian motion and Johnson noise. *Am. J. Phys.* **64**, 225–240 (1996).
15. Gimzewski, J. K., Gerber, Ch., Meyer, E. & Schlitter, R. R. Observation of a chemical reaction using a micromechanical sensor. *Chem. Phys. Lett.* **217**, 589–594 (1994).
16. Barnes, J. R., Stephenson, R. J., Welland, M. E., Gerber, Ch. & Gimzewski, J. K. Photothermal spectroscopy with femtojoule sensitivity using a micromechanical device. *Nature* **372**, 79–81 (1994).
17. Datskos, P. G., Lavrik, N. V. & Rajic, S. Performance of uncooled microcantilever thermal detectors. *Rev. Sci. Instrum.* **75**, 1134–1148 (2004).
18. Yang, J., Ono, T. & Esashi, M. Surface effects and high quality factor in ultrathin single-crystal silicon cantilever. *Appl. Phys. Lett.* **77**, 3860–3862 (2000).
19. Pai, A., Dhurandhar, S. V., Hello, P. & Vinet, J. Y. Radiation pressure induced instabilities in laser interferometric detectors of gravitational waves. *Eur. Phys. J. D* **8**, 333–346 (2000).

Acknowledgements We thank T. Hänsch, A. Imamoglu, R. Warburton and S. Huant for discussions. The Deutsche Forschungsgemeinschaft (DFG) funded this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.K. (karrai@imu.de).

Plasma devices to guide and collimate a high density of MeV electrons

R. Kodama¹, Y. Sentoku², Z. L. Chen¹, G. R. Kumar^{1*}, S. P. Hatchett³, Y. Toyama¹, T. E. Cowan², R. R. Freeman⁴, J. Fuchs^{2*}, Y. Izawa¹, M. H. Key³, Y. Kitagawa¹, K. Kondo⁵, T. Matsuoka¹, H. Nakamura¹, M. Nakatsutsumi¹, P. A. Norreys⁶, T. Norimatsu¹, R. A. Snavely³, R. B. Stephens⁷, M. Tampo¹, K. A. Tanaka⁵ & T. Yabuuchi¹

¹Institute of Laser Engineering, Osaka University, 2-6 Yamada-oka, Suita Osaka 565-0871, Japan

²University of Nevada, Department of Physics, MS-220, Reno, Nevada 89557, USA

³University of California, Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94550, USA

⁴The Ohio State University, Columbus, Ohio 43210, and University of California, Davis, Department of Applied Science, Livermore, California 94550, USA

⁵Faculty of Engineering and Institute of Laser Engineering Osaka University, 2-6 Yamada-oka, Suita Osaka 565-0871, Japan

⁶Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, OX11 0QX, UK

⁷General Atomics, San Diego, California 92186, USA

* Permanent addresses: Tata Institute of Fundamental Research, Mumbai 400 005, India (G.R.K.); Laboratoire pour l'Utilisation des Lasers Intenses, UMR 7605 CNRS-CEA-École Polytechnique-Univ. Paris VI, 91128 Palaiseau, France (J.F.)

The development of ultra-intense lasers¹ has facilitated new studies in laboratory astrophysics² and high-density nuclear science³, including laser fusion^{4–7}. Such research relies on the efficient generation of enormous numbers of high-energy charged particles. For example, laser–matter interactions

at petawatt (10^{15} W) power levels can create pulses of MeV electrons^{8–10} with current densities as large as 10^{12} A cm^{−2}. However, the divergence of these particle beams⁵ usually reduces the current density to a few times 10^6 A cm^{−2} at distances of the order of centimetres from the source. The invention of devices that can direct such intense, pulsed energetic beams will revolutionize their applications. Here we report high-conductivity devices consisting of transient plasmas that increase the energy density of MeV electrons generated in laser–matter interactions by more than one order of magnitude. A plasma fibre created on a hollow-cone target guides and collimates electrons in a manner akin to the control of light by an optical fibre and collimator. Such plasma devices hold promise for applications using high energy-density particles and should trigger growth in charged particle optics.

Electron beams with enormous energy density can be guided by a high-conductivity device consisting of transient high-density plasmas, as the forward-directed energetic electron current can be compensated by a high-density return current in these plasmas. Shaped-plasma devices, such as the hollow-cone and fine fibre-like plasmas shown in Fig. 1a, guide and collimate the high-density energetic-electron beam generated by ultra-intense laser light. Figure 1a shows the electric field contours in the shaped plasma at different times, as generated by a two-dimensional particle-in-cell (PIC) simulation¹¹, showing the electron propagation being accompanied by the fields. Ultra-intense laser light (8×10^{18} W cm^{−2}) is injected into the hollow cone, and generates high-density MeV electrons with an average energy of 3.3 MeV at the tip of the cone. These electrons propagate along the hollow cone's interior surface¹¹ and in the fibre-like plasma. The large current is compensated by the return current set up by the high-density electrons inside the bulk of the fibre.

Propagation (with an angular spread) of the high-density MeV electrons outside the fibre-like plasma induces a strong radial electric field (E_y ; TV m^{−1}) surrounding the fibre-like plasmas (Fig. 1b), and acts to pinch the MeV electrons along the fibre length. A strong azimuthal magnetic field (B_z) of the order of a few hundred MG is also created, and this diverts the MeV electron propagation. The maximum E_y and B_z are given by $E_y = 4\pi eN_e$ and $B_z = 4\pi eN_e(v_e/c)$, respectively. Here N_e is number of electrons surrounding the wire, and v_e is the hot electrons' mean velocity. When the electron velocity is close to the speed of light c , the electric force eE_y and the magnetic force $e(v_e/c)B_z \approx eB_z$ are balanced. Therefore the fast electrons, which are pushed outside by the magnetic field, are returned into the wire by the sheath fields. This field balance results in the collimation and confinement of the electrons in the fibre-like plasmas. Such MeV electron guiding and collimation in the hollow-cone¹¹ and fibre-like plasmas are clearly seen in the electron trace images shown in Fig. 1c. The electrons propagate along the wire with an oscillation due to the surrounding electric and magnetic fields. During the propagation, the transverse emittance of the beam is also reduced by a loss of its transverse energy to the bulk plasma and/or energetic ions via the radial electric field, resulting in beam cooling, which may be analogous to mode selection of light with a single-mode fibre.

We now experimentally demonstrate guiding and collimation of the high-density MeV electrons in the fibre-like plasma device by using a fine carbon wire (5 μ m diameter and 1 mm length) attached to a hollow-cone target⁵ (Fig. 2a). The size of the target in the experiment is not comparable to the condition in the simulation code. However, the guiding and collimation of the electrons predicted by the simulation is qualitatively proved by using the target shown in Fig. 2a in the experiments. Laser light (250–300 TW, 0.6–0.8 ps) with an energy of 180 J (ref. 12) was injected into the cone and focused at the tip of the cone. The hollow-cone side wall can guide the laser light and electrons to the 30- μ m tip of the cone, resulting in the enhancement of the MeV electron flux by a factor of

about 2–3 as compared with that in a simple plane target, as observed in different experiments¹³. Then the energy of the high-density MeV electrons could be about 70 J at the tip of the cone (40% coupling from the laser compared with about 20% for the plane target case¹⁰). The electrons generated at the tip of the cone propagate in the fibre-like plasmas along the wire axis.

Figure 2b shows the >3.5 MeV electron beam contour from the cone wire. The divergence of the electron beam from the wire is about 5° , much smaller than the $20\text{--}30^\circ$ from the cone without the wire. In contrast, a plane target has a divergence of $30\text{--}40^\circ$. The peak number density of the electrons (>3.5 MeV) from the straight wire is also enhanced, being $(0.9\text{--}1.1) \times 10^{12} \text{ sr}^{-1}$ as compared with $1 \times 10^{11} \text{ sr}^{-1}$ from the cone without the wire. The relative energy flux of the electrons in the fine wire attached to the cone could be significantly higher than that in a simple plane target (by a factor of

$20\text{--}30$). Deflection of these high-density MeV electrons with the fine wire is proved by tilting the wire—the emission peak appears on the tilted wire axis— 15° from the cone axis, as shown in Fig. 2c. These experimental results indicate a striking analogy—that the high-density MeV electrons can be guided and collimated by the cone-attached fine wire target in a manner analogous to photon transport in an optical fibre and collimator.

The fine wire is instantaneously heated by bulk-electron return currents inside (compensating the current due to the high-density MeV electron propagation) and becomes a fibre-like plasma. The plasma expansion in the direction perpendicular to the wire axis, which happens long after the propagation of the energetic electrons (that travel at a speed close to that of light), is observed with a two-dimensional (2D) spatially resolved optical high-speed sampling camera¹⁴ in the wavelength range 400–600 nm with a time resolution of 45 ps. For the target with a tilted wire, the plasma expands along the wire in the direction perpendicular to the 15° tilted axis. In Fig. 3a, we show the expansion profiles at different times of the plasmas in the direction perpendicular to the wire axis on the optical images, and we compare them with the profiles predicted

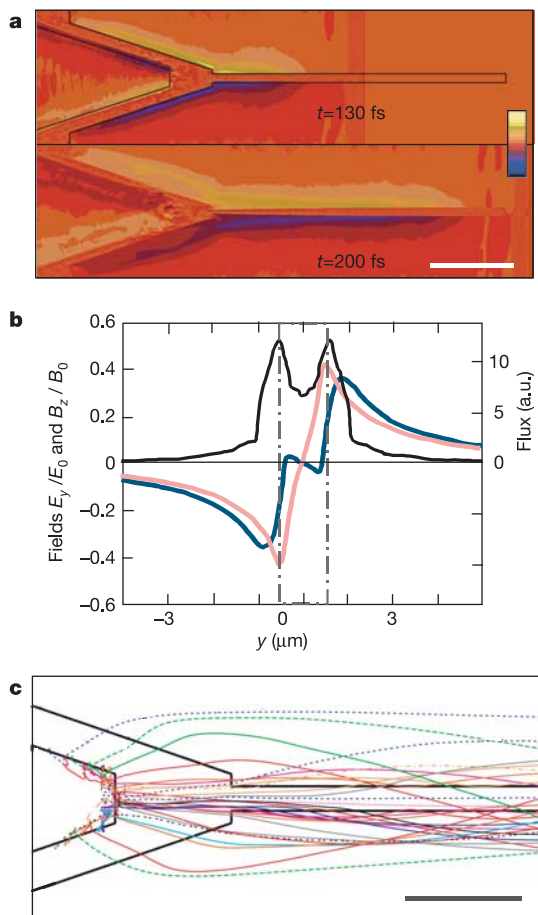


Figure 1 Two-dimensional cartesian particle-in-cell (PIC) simulation modelling the device consisting of hollow-cone and fine fibre-like plasmas. These simulations demonstrate the principle of guiding and collimation of high-density energetic electrons generated by ultra-intense laser light ($8 \times 10^{18} \text{ W cm}^{-2}$). Collisionless transport is assumed to illustrate the guiding, which is related to the surrounding fields induced by the current due to the collisionless MeV electrons on the wire surface. **a**, Initial plasma structure and electric field contours at different times (130 fs and 200 fs). The white scale bar corresponds to $10 \mu\text{m}$, and the colour bar shows the colour coding for electric fields between $-6 \times 10^{12} \text{ V m}^{-1}$ and $6 \times 10^{12} \text{ V m}^{-1}$. The plasma is initially fully ionized and has an electron density of 10^{22} cm^{-3} . **b**, Radial electric field profile (E_y ; blue line) and azimuthal magnetic field (B_z ; red line) around the wire, and energy flux of forward-going electrons (green line), in the direction perpendicular to the wire axis at $20 \mu\text{m}$ from the tip of the cone. The fields are normalized to the incident laser ($E_0 = 9.6 \times 10^{12} \text{ V m}^{-1}$, $B_0 = 320 \text{ MG}$), and the energy flux is in arbitrary units (a.u.) of $10^{18} \text{ W cm}^{-2}$. **c**, Typical trajectories of the MeV electrons, showing the guiding and collimation of the electrons with this plasma device. Black bar, $10 \mu\text{m}$.

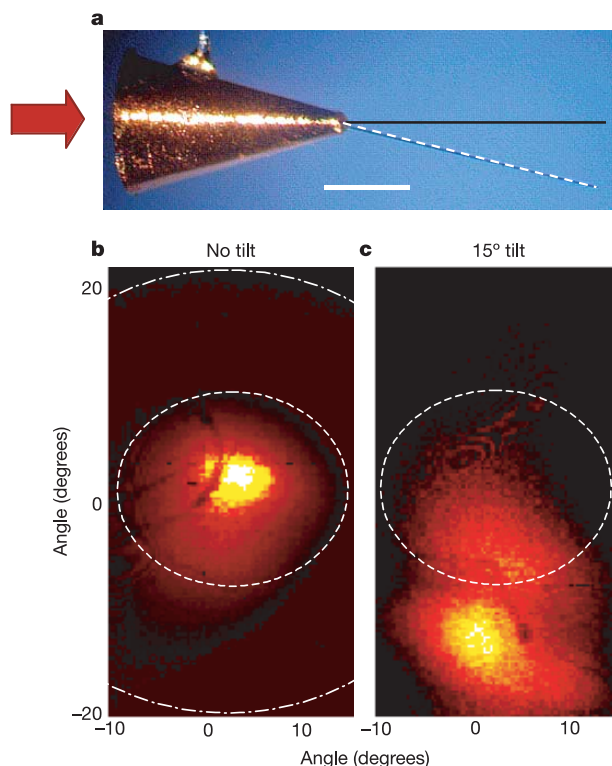


Figure 2 Demonstration of guiding and collimation of high-density MeV electrons generated by 0.3 PW laser light using a fine carbon wire attached to a gold hollow-cone target. **a**, Typical target picture of the straight wire attached to the cone target. White scale bar, $300 \mu\text{m}$. The wire diameter is $5 \mu\text{m}$ and its length 1 mm . Two different types of cone-wire targets were examined. One is with the wire settled along the cone axis (as presented by the black line) and the other is with the wire tilted by 15° from the cone axis (as shown by the dashed white line). The laser light is focused on the tip of the hollow cone inside. **b**, Spatial distribution image of >3.5 MeV electrons from the straight wire target. Six sheets of imaging plates (IPs)¹⁷ are placed at 3 cm from the wire end to measure the electron divergence with energies of $>1 \text{ MeV}$ to $>20 \text{ eV}$. The energy range for each IP is determined taking account of the total thickness of 2 mm Cu foils set between IPs, proton track detectors (CR-39) placed in front of the stacked IPs and the IPs in front of a specific IP. The dot-dashed circle shows the electron beam divergence (FWHM) from the plane target, and the dashed circle that from the cone without the wire. The divergence (FWHM) on the contours roughly corresponds to the yellow region for the wire target. **c**, The IP image of >3.5 MeV electrons from the 15° tilted wire target. The dashed circle shows the FWHM image of the electrons when there is no wire target.

by a one-dimensional hydrodynamic simulation (LASNEX)¹⁵ with cylindrical symmetry. The simulation profiles consistently reproduce the experiment with temperatures of 2–4 keV. The radius of the profile corresponding to the 10% points of the maximum intensity from the simulation is plotted for those in the different ‘snapshots’ of the sampling camera as shown in Fig. 3b, that is, at times 45 ps, 90 ps and 135 ps.

The expansion rate from the simulation clearly depends on the temperature of the heated wire, and the experiment is quite consistent with the calculation for temperatures of 2–4 keV. If there was no rapid expansion of the wire due to a static field surrounding it and the fibre was immediately heated up to 3 keV due to the MeV electron propagation, pressures of 3 Gbar would be expected for fully ionized solid carbon fibre plasmas. Simple estimates based on energy conservation (ignoring all the energy loss of radiation and energetic ions via strong fields) and classical stopping power of the energetic electrons with a slope temperature of 2 MeV (for a maxwellian distribution) imply that at least 6% (11 J) of the laser energy would be required for the electrons

propagating in the fibre to create such plasmas. Taking account of the coupling from the laser to MeV electrons in the cone geometry (about 40% or 70 J)^{10,13}, the fibre-like plasma attached to the hollow cone could guide at least 15% (11 J) of the total MeV electrons (70 J) generated by the ultra-intense laser light. On the other hand, the 5 μm fibre might geometrically select only <2.8% of the total electron emission from the 30 μm tip of the cone, which is less than the experimental value by one order of magnitude. Therefore the fibre-like plasma created on the cone tip efficiently guides and collimates the high-density MeV electrons as predicted by the PIC simulation, resulting in an extremely high energy-density plasma state, possibly with gigabar pressures—and all this with only 180 J laser pulse energy delivered to target.

Collimation of the energetic electrons with different shaped plasma devices to increase the energy density was also examined

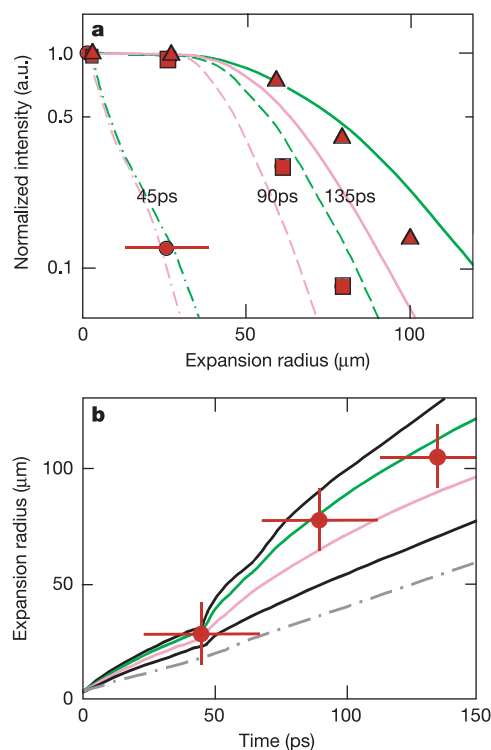


Figure 3 Expansion of the fibre-like plasma heated by high-density energetic electrons propagating along the wire. **a**, The expansion profiles of the plasmas in the direction perpendicular to the wire axis obtained with the sampling camera (400–600 nm) at different times (circles: 45 ps; squares: 90 ps; triangles: 135 ps) and the profiles (dot-dashed line: 45 ps; dashed line: 90 ps; solid line: 135 ps) predicted by the hydrodynamic simulation (LASNEX). The error bars correspond to the spatial and time resolutions of the diagnostics. The profiles are normalized to the peak intensity. The green and red lines show 2 keV and 4 keV as an initial heated temperature. In the simulation, the 5 μm carbon wire is assumed to be instantaneously heated to a uniform initial temperature at time 0, and is then allowed to expand and cool with no further energy deposition and no field effect on the expansion. Radiation transport is handled with multi-group diffusion, with five groups in particular in the 400–600 nm band pass assumed for the sampling camera. To simulate the sampling camera’s response, the instantaneous intensity is integrated over the 45 ps. **b**, The radius of the profile corresponding to the 10% points of the maximum intensity from the simulation for different heating temperatures (dot-dashed line: 0.5 keV; grey: 1 keV; red: 2 keV; green: 4 keV; blue: 6 keV) and from the experiments at different frame times (filled circles). The error bars correspond to the spatial and time resolutions of the diagnostics.

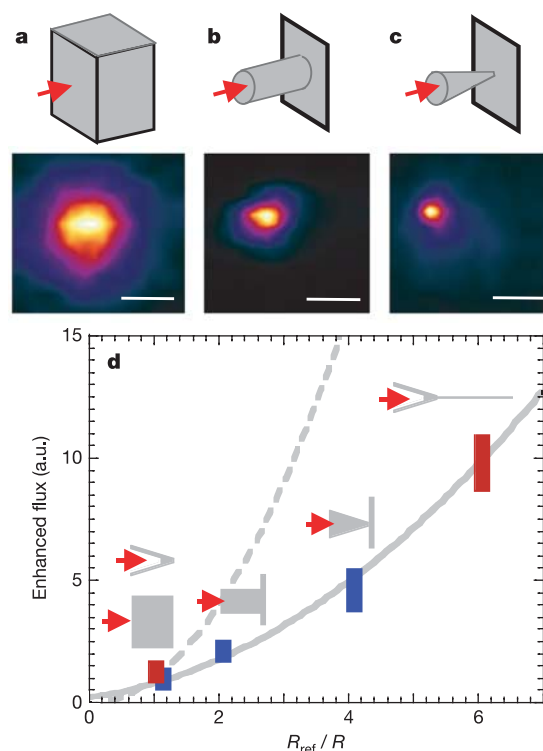


Figure 4 Collimation of MeV electrons in the shaped plasma conductors and enhancement of the electron peak flux with the collimators. **a–c**, Bottom row, the optical emission (400–600 nm) from the rear side of the 10 μm Al plane attached to the different shaped conductors (top row) obtained with the high-speed sampling camera. The emission is temporally separated from any other hydrodynamic heating. **a**, 200- μm -thick Al block; **b**, 200- μm -long Al cylinder with 100 μm diameter attached to the Al foil; **c**, 30° solid Al cone with a length of 200 μm and a tip size of about 40 μm attached to the Al foil. The colour contour in the optical emission data is normalized to the peak intensity for each of the targets. White scale bars correspond to 200 μm on the target. **d**, Relative peak flux of the collimated electrons for different shaped targets (two data sets) as a function of R_{ref}/R ; here R is the radius of the emission source, and R_{ref} is the radius of a reference source, used for normalization. The first data set consists of the plane block, the cylinder and the taper tip (obtained with the visible rear-side emission, **a–c**). This data set is plotted as blue squares and normalized by the plane data (reference of the first data set). The second data set consists of the hollow cone and the cone–wire, and is obtained with the IPs. This data set is shown as red squares normalized by the cone data (reference of the second data set). Original tip sizes of the targets are used as the source size for the second set—that is, 30 μm diameter for the cone and 5 μm diameter for the wire. The peak flux for the hollow cone (second reference) is about 2–3 times larger than that for the plane target (first reference)¹³. The dotted line shows the enhancement according to the simple geometrical optics determined by the source size or $1/R$ and the solid line shows the experimental scaling.

using shaped Al targets (Fig. 4a–c). The relative energy flux in an Al solid cylinder (length 200 μm , diameter 100 μm) and an Al solid taper (length 200 μm ; taper angle 30°; the tip or attached point size, 30–40 μm) attached to a 10 μm Al plate was measured and compared with that in a 200- μm -thick planar target. High-density MeV electrons generated at the flat surface propagated through the shaped target to the plane target. Optical emission due to the energetic electrons from the rear side of the plane Al target was recorded with the 2D high-speed sampling camera. Good, shape-dependent collimation of the electrons is clearly seen on the emission images of the Al plate rear side (Fig. 4a–c). The peak flux of the emission was enhanced in the cylinder by a factor of about 2, and in the taper by a factor of 4.5, as compared with that in the block target. The relative peak energy fluxes for the different shaped targets (cylinder and taper tip), normalized to that for the plane block, are plotted in Fig. 4d as a function of the reciprocal of the emission source size $1/R$. The enhancement factor is less than that estimated from simple geometrical optics (as given by the source size) but still indicates that such plasma devices guide and collimate the high-density electrons in a similar manner to the control of light with a taper-shaped optic-block collimator.

The enhancement of the peak energy flux of the electrons for the cone-wire target obtained with imaging plates is also plotted in Fig. 4d as compared to that from the cone target without the wire. The cone-wire data (normalized by the hollow-cone case) shows a scaling similar to the data set of the cylinder and the taper tip (normalized by the plane target case), while the peak flux for the hollow cone is about 2–3 times larger than that for the plane target¹³. The enhancement for the cone-wire is also less than the expectation from simple geometrical optics given by the source size difference. These differences could be due to the escape of higher energy electrons from the shaped target and the wire overcoming the strong surrounding fields. About 10% of the energy should also be lost in the long wire, from a simple estimation for classical stopping. Even with these losses, the electron energy density for the cone-wire target can be higher than that for the simple planar target by a factor of about 20–30, that is, 2–3 times higher with the hollow cone and 10 times higher with the wire. These experimental results prove that the high-density MeV electrons can be guided and collimated with the plasma devices in a manner analogous to light control with a taper-shaped block and a fibre optic collimator, resulting in an extremely high energy density plasma state.

These devices to guide and control the high energy-density particles are free from any damage threshold of the bulk modulus of the solid-state material because of the fast transient phenomena. Plasmas, which are created by fast heating caused by high-density energetic electrons in ultra-intense laser plasma interactions, can be shaped in a unique way to control charged particles at high flux (TA cm^{-2}) with an enormous energy density in a fashion analogous to light control with a conventional optical device. It might be appropriate to call them ‘plasma photonic devices’, and we expect that they could be used in many future applications because of the control of the high energy-density state allowed by delivery of the laser energy and shaping of the target. Some examples are: point sources of ultra-intense pulse radiation for medical and material science applications; higher energy-density plasma with a higher field for laboratory astrophysics; isochoric heating for precise estimation of equations of state; higher temperature and longer, yet finer plasma for a harder X-ray laser amplification medium¹⁶; and high-density MeV electron guiding for an ultra-compact free electron laser. □

Received 8 July; accepted 21 October 2004; doi:10.1038/nature03133.

1. Perry, M. D. & Mourou, G. Terawatt to petawatt subpicosecond lasers. *Science* **264**, 917–924 (1994).
2. Remington, B. A., Arnet, D., Drake, R. P. & Takabe, H. Modeling astrophysical phenomena in the laboratory with intense lasers. *Science* **284**, 1488–1493 (1999).
3. Ledingham, K. W. D., McKenna, P. & Singhal, R. P. Applications for nuclear phenomena generated by ultra-intense lasers. *Science* **300**, 1107–1111 (2003).

4. Tabak, M. *et al.* Ignition and high gain with ultra powerful lasers. *Phys. Plasmas* **1**, 1626–1634 (1994).
5. Kodama, R. *et al.* Fast heating of ultrahigh-density plasma as a step towards laser fusion ignition. *Nature* **412**, 798–802 (2001).
6. Key, M. H. Fast track to fusion energy. *Nature* **412**, 775–776 (2001).
7. Kodama, R. *et al.* Fast heating scalable to laser fusion ignition. *Nature* **418**, 933–934 (2002).
8. Wharton, K. B. *et al.* Experimental measurements of hot electrons generated by ultraintense ($>10^{19}\text{W/cm}^2$) laser plasma interactions on solid-density targets. *Phys. Rev. Lett.* **81**, 822–825 (1998).
9. Key, M. H. *et al.* Hot electron production and heating by hot electrons in fast ignitor research. *Phys. Plasmas* **5**, 1966–1972 (1998).
10. Kodama, R. *et al.* Fast ignition research at the Institute of Laser Engineering Osaka University. *Phys. Plasmas* **8**, 2268–2274 (2001).
11. Sentoku, Y. *et al.* Laser light and hot electron micro focusing using a conical target. *Phys. Plasmas* **11**, 3083–3087 (2004).
12. Kitagawa, Y. *et al.* Prepulse-free petawatt laser for a fast ignitor. *IEEE J. Quant. Electron.* **40**, 281–293 (2004).
13. Kodama, R. *et al.* In *Inertial Fusion Science and Application 2003* (eds Hammel, B. A., Meyerhofer, D. D., Meyer-ter-Vehn, J. & Azechi, H.) 333–338 (American Nuclear Society, Illinois, 2004).
14. Kodama, R. *et al.* Development of a two-dimensional space-resolved high speed sampling camera. *Rev. Sci. Instrum.* **70**, 625–628 (1999).
15. Zimmerman, G. B. & Krue, W. L. Numerical simulation of laser-initiated fusion. *Comment Plasma Phys. Controlled Fusion* **2**, 51–61 (1975).
16. Fill, E. *et al.* in *X-ray Lasers 1998* (eds Kato, Y., Takuma, H. & Daido, H.) 301–308 (Inst. Phys. Conf. Ser. No. 159, IOP Publishing, Bristol, 1999).
17. Tanaka, K. A. *et al.* Calibration of imaging plate for high energy electron spectrometer. *Rev. Sci. Instrum.* (in the press).

Acknowledgements We thank the target fabrication, laser operation and acquisition groups at ILE Osaka University. G.R.K. thanks JSPS for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (ryo@ile.osaka-u.ac.jp).

Bismuth embrittlement of copper is an atomic size effect

Rainer Schweinfest*, Anthony T. Paxton & Michael W. Finnis

Atomistic Simulation Centre, Department of Physics and Astronomy, Queen's University Belfast, Belfast BT7 1NN, UK

* Present address: Science+Computing ag, Hagellocher Weg 71–5, 72070 Tübingen, Germany

Embrittlement by the segregation of impurity elements to grain boundaries is one of a small number of phenomena that can lead to metallurgical failure by fast fracture¹. Here we settle a question that has been debated for over a hundred years²: how can minute traces of bismuth in copper cause this ductile metal to fail in a brittle manner? Three hypotheses for Bi embrittlement of Cu exist: two assign an electronic effect to either a strengthening³ or weakening⁴ of bonds, the third postulates a simple atomic size effect⁵. Here we report first principles quantum mechanical calculations that allow us to reject the electronic hypotheses, while supporting a size effect. We show that upon segregation to the grain boundary, the large Bi atoms weaken the interatomic bonding by pushing apart the Cu atoms at the interface. The resolution of the mechanism underlying grain boundary weakening should be relevant for all cases of embrittlement by over-size impurities.

As A. H. Cottrell has put it “when a large force is applied to a crystal two things may happen: the atoms in the crystal may slide past one another; and they may pull apart.” In the former case, a metal will fail gracefully with absorption of energy by dislocation generation and plastic deformation: the metal is said to be tough. Otherwise (in the latter case) the metal will fail by fracture. The applied force to cause failure is governed by the energy release rate, G , which is the energy that may be derived from the crystal plus loading train per unit area of crack advance⁶. In defining G one

Table 1 Elastic constants (GPa) of Cu, Cu₂₆Bi, Cu₇Bi and Cu₇Pb

	<i>K</i>	<i>C'</i>			<i>C</i> ₄₄			μ_t		
		0.011	0.012	0.013	0.011	0.012	0.013	0.011	0.012	0.013
Cu	190	35	25	18	99	74	54	56	41	27
Cu ₂₆ Bi	168	32	25		102	75		55	42	
Cu ₇ Bi	143	29		13	126		60	61		29
Cu ₇ Pb	150	27		16	125		69	60		34

For each formula, the bulk modulus (*K*) is calculated from the second derivative of the energy–volume relationship at the calculated equilibrium volume. The atomic volumes Ω (in units of nm³) 0.012 and 0.013 are the equilibrium volumes found for the two Cu–Bi compositions. The shear constants *C'*, *c*₄₄ and μ_t (see text for definitions) are calculated for each formula at its equilibrium volume and at the equilibrium volume of pure Cu, 0.011 nm³.

imagines releasing workless constraints holding together the faces of unit area of crack under load. As a consequence, the specimen's elastic stiffness is reduced. In the case of fixed-grip loading, *G* is the elastic relaxation energy of the specimen; in dead-weight loading, *G* also contains the work done by the weight as it descends under gravity. Nevertheless, *G* is independent of the loading configuration⁶. In equilibrium, this energy is exactly balanced by the cohesive energy bonding together the two halves of crystal ahead of the crack tip. In the absence of irreversible plastic work, this equates with the excess free energy per unit area γ_s of the two surfaces exposed when the crack advances. Then we have $G_c^{\text{cleav}} = 2\gamma_s$, the critical energy release rate for cleavage; one thinks of this as the crystal's resistance to fracturing or 'work of fracture'⁷. When a crack exists in a crystal, its sharpness induces a concentration of the applied stress just ahead of the tip. If this can be relieved by dislocation nucleation and motion then the tip will be blunted and the component safe from fracture; a critical energy release rate G_c^{disl} may still be defined, and is a complicated function of the stress intensity, elastic constants and crystal defect energies⁸. A ductile-to-brittle transition (DBT) occurs when G_c^{cleav} becomes smaller than G_c^{disl} (ref. 9).

It has been known for over 100 years that copper can be embrittled by concentrations of bismuth smaller than 100 parts per million (ref. 2). The otherwise tough metal can be shown to practically 'fall apart' at the grain boundaries (GBs), exposing faceted crystal faces at the fracture surfaces¹⁰. This system has become a model for the study of segregation embrittlement in a number of carefully conducted experiments and calculations. It is not possible to calculate G_c^{disl} from first principles, but sophisticated estimates using anisotropic elastic theory show that at many GBs in pure Cu, G_c^{disl} is around 1–2 J m^{−2} (ref. 8). We find (see Methods) that for a typical GB, G_c^{cleav} is 3.1 J m^{−2} in pure Cu. Upon segregation of one monolayer of Bi to the same GB, G_c^{cleav} falls to just 1.33 J m^{−2}, which accounts rather well for the observed DBT. It remains to explain why Bi embrittles the GB. The two hypotheses based on electronic structure are (1) that the presence of Bi stiffens (strengthens) the bonds of the surrounding metal, and thereby inhibits the bond mobility that is a necessary condition for plasticity³, and (2) that bonds across the GB are weakened by the presence of the Bi (ref. 4). In simplest terms, the DBT arises from an increase in G_c^{disl} in (1), and a decrease in G_c^{cleav} in (2). We are able to test and reject the first hypothesis. The second we support, but we find that the generally accepted explanation is incorrect. The current view is that the weakening is an electronic effect: the more electro-negative segregant withdraws electrons from the metal *d* bands, and consequently reduces the bond strength. This supposition has retained its currency on the basis of the observation of 'white lines' in the inelastic electron scattering intensity observed at Bi-segregated GBs in the electron microscope. We refute these ideas, and propose instead that the embrittlement is essentially a size effect as earlier suggested by Sutton and Vitek⁵. Two conditions are necessary for GB segregation embrittlement in Cu: a segregant that is virtually insoluble, and that also has a very much larger atomic radius. The segregant then segregates to the boundary, where it literally pushes apart the Cu atoms across the interface and

weakens the interatomic bonding. We have repeated all calculations with Pb in place of Bi to verify these conclusions. This accounts for the embrittlement caused by Pb, Hg and Bi (ref. 11); and indeed we find that Pb and Bi are practically indistinguishable in the properties we have studied and report here, contrary to the expectations of an 'electronic' picture of embrittlement.

First, let us dispose of the question of bond stiffening. Most simply we can do this by calculating stress–strain relations in crystals of Cu containing a certain concentration of Bi or Pb atoms. If the hypothesis were correct, we would expect Bi to increase the elastic moduli and the theoretical strength. Consequent to this would be an increase in G_c^{disl} . The bulk moduli, *K*, of pure Cu, Cu₂₆Bi and Cu₇Bi are respectively 190, 172 and 143 GPa, reflecting an increase in atomic volume Ω due to the insertion of impurity atoms with a large atomic radius. If Bi is replaced with Pb, the values are 190, 176 and 150 GPa. Table 1 shows these data along with calculated shear constants *C'*, *c*₄₄ and $\mu_t = (2C' + c_{44})/3$. The most notable fact is that the shear constant relevant to the emission of a lattice dislocation, μ_t , is completely insensitive to the presence of either Bi or Pb at a given Ω . Furthermore, as shown in Fig. 1, the Frenkel theoretical strength¹² is found similarly to depend only on the atomic volume, and not on the quantity of either Bi or Pb in the lattice. These data point rather clearly to a simple size effect, there being little difference between the effect of Bi and Pb.

An unequivocal pointer to a size effect is contained in the data in Table 1. At first sight, it is curious that at constant volume, *C'* is softened by Bi while *c*₄₄ is hardened. In fact, this is precisely the effect expected were the Cu–Bi bonding to be of a central force type. Suppose the impurity acted as an additional spherically symmetric centre of repulsive potential, which has the value *V*(*R*) at nearest

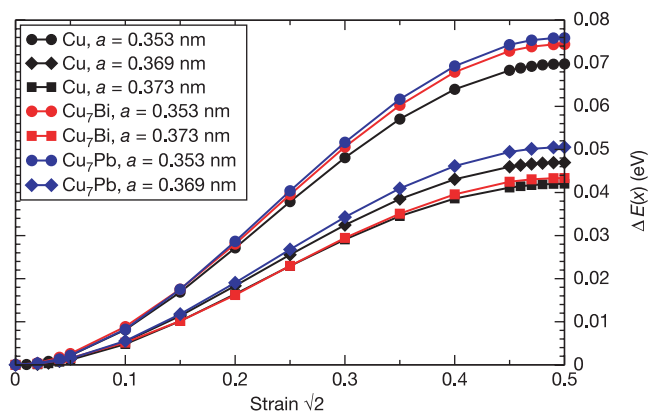


Figure 1 Internal energy as a function of strain as a unit cell based on a face-centred cubic lattice is sheared into itself on a (111) plane in a [112] crystal direction¹². This is a measure of the response of the crystal to large deformations of the type that are required when a crack tip is blunted by the emission of a dislocation²⁷. Calculations are made at the equilibrium lattice constants *a* found for each composition (Table 1). The maximum slope and height of these curves and the curvature around the origin, proportional to the shear constant μ_t , are remarkably insensitive to impurity at a given average atomic volume (Table 1).

neighbours. We have extracted values for its derivatives V' (the Kanzaki forces¹³) and V'' by making calculations of forces in pure copper at all the same atomic configurations that appear during the calculations of shear constants with a Bi atom present. Remarkably, the results are insensitive to Bi concentration. Furthermore, the harmonic theory for a nearest-neighbour face-centred cubic lattice¹⁴ shows that such a potential will harden c_{44} and soften C' if $3 < RV''/V' < 7$. Our calculated value of this critical ratio, namely 5.3–5.6, is comfortably within the range 3–7. This is very strong evidence of a simple size effect.

Second, we use explicit calculation to account for the reduction in cohesion at an embrittled GB. We study the $[\bar{1}10](331)$ symmetric tilt GB in Cu (sometimes designated $\Sigma 19a$), which is known to be embrittled by Bi (see Fig. 2). The critical energy release rate should be written $G_c^{\text{cleav}} = W_{\text{sep}} + W_p$ (refs 9, 15). Here W_{sep} is the area under the force–displacement curve as the GB is cleaved, with no energy dissipation through heating or plasticity, at constant composition with the GB impurity evenly distributed between the two exposed surfaces¹⁶. W_p is the energy per unit area associated with irreversible plastic work during crack advance. Although $W_p \gg W_{\text{sep}}$, it is admissible to consider only the first term: firstly because a weakened interface has a reduced stress intensity factor and larger decohesion zone⁹, and secondly, W_p is a monotonic function of W_{sep} and so if W_{sep} goes to zero, so does W_p (ref. 15). Hence we examine the behaviour of $W_{\text{sep}} = 2\gamma_s - \gamma_{\text{gb}}$, in which γ_{gb} is the GB energy (with or without segregant) and γ_s is the surface energy—in this case of the (331) surface of Cu (again with or without segregant). The thermodynamic excess concentration of segregant at the GB, Γ , can be estimated from the bulk concentration and the segregation energy, E_{seg} , defined as the difference in internal energy at $T = 0$ K between a Bi atom at a GB site and one at a bulk site. We find that the work of separation of the $\Sigma 19a$ GB is reduced from 3.1 J m^{-2} in pure Cu to 1.33 J m^{-2} in the presence of a monolayer (ML), Fig. 2, of Bi at the GB. Results are presented in Table 2. E_{seg} (we neglect vibrational entropy) is dependent on the GB concentration; however, the values are so large that, based on the McLean adsorption isotherm¹⁷, all available GB sites will be occupied by an impurity. This is consistent with the observation that the maximum boundary concentration is 22 atoms nm^{-2} , equivalent to 1.2 ML of a Cu (111) plane¹⁸. The resultant decohesion is very evident in our value of W_{sep} , and is a simple consequence of an excess of volume at the GB plane.

Third, we can eliminate the hypothesis that the effect is electronic

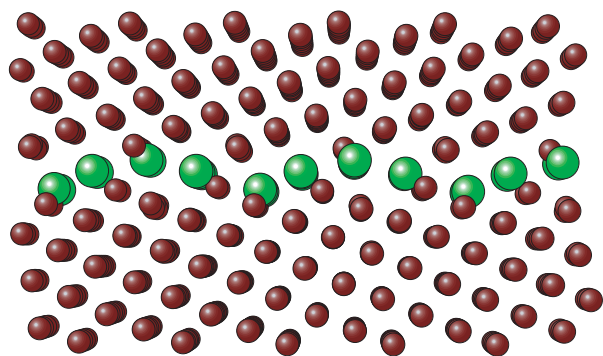


Figure 2 Structure of a 1 ML Bi-segregated $\Sigma 19a$ grain boundary in Cu. We have incorporated the Bi atoms over 1, 2 and 3 layers of Cu atoms adjacent to the GB plane. The lowest energy 1 ML structures, in the absence of additional vacancies, are those in which the Bi is distributed over two GB planes. However, we show here a configuration which is just $0.007 \text{ eV atom}^{-1}$ (0.17 J m^{-2}) higher in energy, in which three Cu layers are decorated with Bi impurities. The resulting relaxed structure is striking, because it displays the incipient formation of GB faceting which is observed at the Bi-segregated $\Sigma 19a$ GB in Cu (ref. 10). Red spheres are Cu; green are Bi.

Table 2 Work of separation and segregation energy of $\Sigma 19a$ GB in Cu

Parameter	Pure Cu	Cu–Bi		Cu–Pb	
Γ		1 ML	0.5 ML	1 ML	0.5 ML
$-E_{\text{seg}}$ (eV)		1.58	1.32	1.47	1.40
γ_{gb} (J m^{-2})	1.04				
W_{sep} (J m^{-2})	3.10	1.33	2.07	1.59	2.21

See text for definition of parameters.

in the sense that the decohesion arises from the ‘draining’ of electrons from the Cu–Cu bonds by the more electronegative impurity atom. This commonly held view arises from early work⁴ that omitted considerations of GB geometry, metallic screening or the evaluation of total energy¹⁹. The view has been supported by experiments searching for evidence of holes in the Cu d band localized at a GB via their signature ‘white line’ in the $L_{2,3}$ absorption edges in spatially resolved electron energy-loss microscopy. These were observed by some authors²⁰, but were not found by others¹⁸. Later, Muller²¹ cast doubt on the former interpretation by showing that their experimental findings could equally well be attributed to core level shifts at the GBs. We explicitly demonstrate the lack of charge transfer by plotting in Fig. 3 the local density of states associated with a Cu atom at the $\Sigma 19a$ GB containing 1 ML Bi. It is very clear that the Fermi level falls about 1.5 eV above the edge of the Cu d band—there are no holes and hence no white line is to be expected. This is consistent with the findings of Alber¹⁸. In fact, the effect of the Bi at the GB is to narrow the d band significantly, which is a consequence of the increase in atomic volume at the grain boundary—a point which lies at the heart of our picture of the embrittlement as a size effect. If anything, the d -electron character at the Fermi level is lessened by the presence of Bi (ref. 22). Although the picture of band bending at segregated metal GBs is a common viewpoint and has been proposed also in the reverse sense²³ (for example, transfer of charge from S to Fe), as has been pointed out earlier²⁴ it is clear that the perfect screening of the bulk metal will neutralize any such effect.

Recent work²² appears to contradict our findings: the $[001](310)$ ($\Sigma 5$) symmetric tilt GB was found to have almost no expansion. These authors attribute embrittlement to an electronic effect whereby the Cu d band retreats below the Fermi level, reducing sd hybridization and rendering the metal ‘zinc-like’. Unfortunately, the boundary was not demonstrated to be embrittled either in the calculations or in the measurements, and indeed their observed

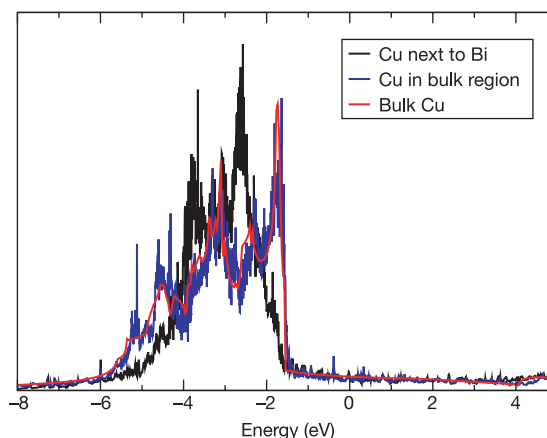


Figure 3 The local density of states (in arbitrary units) projected onto a Cu atom in a $\Sigma 19a$ GB to which 1 ML of Bi is segregated, the $\Sigma 19a$ in pure Cu and pure bulk copper. The Fermi energy is at 0 eV, and neglecting the noisy spikes in the GB curves as due to the very small Brillouin zone involved, it is clear that in all cases the Fermi level falls at least 1.5 eV above the edge of the Cu d band: no white line is predicted in a core level spectrum.

segregated Bi concentration of 2.9 atoms nm⁻² is some 5 times smaller than generally measured at fractured GBs^{18,25}. The electronic effect these authors measure is rather small and cannot account for the presumed reduction in fracture toughness, as there is no reason to suppose that Zn in the crystal structure of Cu will be brittle. They make the point that if a size effect existed, one would expect Cu to be embrittled by Ag. However, Ag, unlike Bi, is soluble in Cu and its atomic radius is 7% smaller. Nonetheless, their comparison is revealing. Although Ag will segregate to GB sites, it need not be forced into one or more monolayers because it may dissolve in the surrounding grains. In the case of insoluble impurities, one could view the embrittlement as the crystal's attempt to expel the impurity to the surface. A full understanding can only be reached by considering equally the grain boundary and surface energies.

We therefore reject the conventional wisdom that embrittlement is caused by charge transfer between the metal atoms and the segregant atoms at the GB. It is often overlooked that Cu is embrittled by Hg and Pb as well as by Bi (ref. 11). Our conclusions are these: Hg, Pb and Bi will segregate to most GBs because of their extremely low solid solubility in the bulk. This itself is essentially a size effect, as many GBs present sites that can accommodate a large impurity atom. The GBs are then embrittled when after segregation $G_c^{\text{cleav}} < G_c^{\text{disl}}$. To be precise, we have actually found $G_c^{\text{cleav}} \gg G_c^{\text{disl}}$ before segregation, and $G_c^{\text{cleav}} \approx G_c^{\text{disl}}$ after segregation. We believe therefore that segregation places Cu close to the ductile-to-brittle transition. This explains why metallurgical treatments can promote or inhibit impurity embrittlement by Bi and Pb. We believe this will prove to be a general result in the case of an oversized impurity in a metal. □

Methods

The questions raised here are ones that are very readily answered using the well known local density approximation to density functional theory. We have used the Vienna *ab initio* Simulation Program, VASP²⁶, using the standard ultrasoft pseudopotentials supplied with the program and a plane-wave cut-off of 270 eV. Brillouin zone integration was done using a modified tetrahedron method and k-meshes equivalent to 14 × 14 × 14 in the primitive face-centred cubic (f.c.c.) unit cell. The pseudopotential for Cu was tested by calculating energies of competing close-packed phases and elastic constants: the lattice constant is correct to within 2%, the bulk modulus within 37%. The Bi pseudopotential was shown to predict the A7 crystal structure of pure Bi (lattice constants within 4%, bulk modulus within 30%) and its transition to body-centred cubic under pressure. The pseudopotential for lead predicts the f.c.c. lattice constant to within 1.4% and the bulk modulus to within 13%. Also the f.c.c. phase is favoured over diamond cubic, which is known to be a relativistic effect of the mass velocity shift of the s electrons near the heavy nucleus.

We have computed the structure and energy of 0.5 ML and 1 ML of Bi and Pb at the Σ19a GB in a large number of configurations in 'supercells' containing a grain boundary and two remote surfaces. Surface energies were determined in separate calculations without a GB present. All energies reported here are obtained from supercells that have been relaxed using molecular statics, residual forces not exceeding 0.01 eV Å⁻¹. Surfaces and GBs were relaxed under conditions of fixed concentrations Γ of impurity. In calculating W_{sep} , the impurity was distributed evenly between the two surfaces exposed by cleavage, in accordance with observation²⁵.

Received 29 June; accepted 5 November 2004; doi:10.1038/nature03198.

- Gray, J. L. Investigation into the consequences of the failure of a turbine-generator at Hinkley Point 'A' power station. *Proc. Inst. Mech. Eng.* **186**, 379–390 (1972).
- Hampe, W. Beiträge zu der Metallurgie des Kupfers. *Berg- u. Hüttenwesen* **23**, 93–137 (1874).
- Haydock, R. The mobility of bonds at metal surfaces. *J. Phys. C* **14**, 3807–3816 (1981).
- Messmer, R. & Briant, C. L. The role of chemical bonding in grain boundary embrittlement. *Acta Metall.* **30**, 457–467 (1982).
- Sutton, A. P. & Vitek, V. An atomistic study of tilt grain boundaries with substitutional impurities. *Acta Metall.* **30**, 2011–2033 (1982).
- Lawn, B. R. *Fracture of Brittle solids* Sect. 2.2, 2nd edn (Cambridge Univ. Press, Cambridge, 1993).
- Kelly, A. & Macmillan, N. H. *Strong Solids* Sect. 2.3.1 (Clarendon, Oxford, 1986).
- Anderson, P. M. & Rice, J. R. Dislocation emission from cracks in crystals or along crystal interfaces. *Scripta Metall.* **20**, 1467–1472 (1986).
- Rice, J. R. & Wang, J.-S. Embrittlement of interfaces by solute segregation. *Mater. Sci. Eng. A* **107**, 23–40 (1989).
- Sigle, W., Chang, L.-S. & Gust, W. On the correlation between grain-boundary segregation, faceting and embrittlement in Bi-doped Cu. *Phil. Mag. A* **82**, 1595–1608 (2002).
- Warke, W. R. in *ASM Handbook* Vol. 11, *Failure Analysis and Prevention* 861–867 (ASM International, Ohio, 2002).
- Paxton, A. T., Gumbsch, P. & Methfessel, M. A quantum mechanical calculation of the theoretical strength of metals. *Phil. Mag. Lett.* **63**, 267–274 (1991).
- Kanzaki, H. Point defects in face-centred cubic lattice—I distortion around defects. *J. Phys. Chem. Solids* **2**, 24–36 (1957).

- Finnis, M. W. The energy and elastic constants of simple metals in terms of pairwise interactions. *J. Phys. F* **4**, 1645–1656 (1974).
- Jokl, M. L., Vitek, V. & McMahon, C. J. Jr A microscopic theory of brittle fracture in deformable solids: a relation between ideal work to fracture and plastic work. *Acta Metall.* **28**, 1479–1788 (1980).
- Finnis, M. W. The theory of metal–ceramic interfaces. *J. Phys. Condens. Matter* **8**, 5811–5836 (1996).
- Sutton, A. P. & Balluffi, R. W. *Interfaces in Crystalline Materials* Ch. 7 (Clarendon, Oxford, 1995).
- Alber, U., Mülleijans, H. & Rühle, M. Bismuth segregation at copper grain-boundaries. *Acta Mater.* **47**, 4047–4060 (1999).
- Goodwin, L., Needs, R. J. & Heine, V. Effect of impurity bonding on grain-boundary embrittlement. *Phys. Rev. Lett.* **60**, 2050–2053 (1988).
- Bruley, J., Keast, V. J. & Williams, D. B. An EELS study of segregation-induced grain-boundary embrittlement of copper. *Acta Mater.* **47**, 4009–4017 (1999).
- Muller, D. A. Why changes in bond lengths and cohesion lead to core-level shifts in metals, and consequences for the spatial difference method. *Ultramicroscopy* **78**, 163–174 (1999).
- Duscher, G., Chisholm, M., Alber, U. & Rühle, M. Bismuth-induced embrittlement of copper grain boundaries. *Nature Mater.* **3**, 621–626 (2004).
- Özkaya, D., Yuan, J., Brown, L. M. & Flewitt, P. E. J. Segregation-induced hole drilling at grain-boundaries. *J. Microsc.* **180**, 300–306 (1995).
- Saqi, M. A. S. & Pettifor, D. G. Role of impurity elements in metal–metal bond strengths. *Phil. Mag. Lett.* **56**, 245–249 (1987).
- Powell, B. D. & Mykura, H. The segregation of bismuth to grain boundaries in copper–bismuth alloys. *Acta Metall.* **21**, 1151–1156 (1973).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab-initio* total energy calculations using a plane wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
- Rice, J. R. Dislocation nucleation from a crack tip: an analysis based on the Peierls concept. *J. Mech. Phys. Solids* **40**, 239–271 (1992).

Acknowledgements Financial support was provided by EPSRC.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.T.P. (Tony.Paxton@QUB.ac.uk).

Meteoric smoke fallout over the Holocene epoch revealed by iridium and platinum in Greenland ice

Paolo Gabrielli^{1,2}, Carlo Barbante^{2,3}, John M. C. Plane⁴, Anita Varga², Sungmin Hong⁵, Giulio Cozzi², Vania Gaspari², Frédéric A. M. Planchon², Warren Cairns⁵, Christophe Ferrari^{1,6}, Paul Crutzen⁷, Paolo Cescon^{2,3} & Claude F. Boutron^{1,8}

¹Laboratoire de Glaciologie et Géophysique de l'Environnement (UMR CNRS/ Université Joseph Fourier 5183), 54, rue Molière, BP 96, 38402 St Martin d'Heres cedex, France

²Department of Environmental Sciences, ³Institute for the Dynamics of Environmental Processes-CNR, University of Venice, Ca' Foscari, 30123 Venice, Italy

⁴School of Environmental Sciences, University of East Anglia, Norwich NR47TJ, UK

⁵Korea Polar Research Institute, Korea Ocean Research & Development Institute, Ansan, PO Box 29, Seoul 425-600, Korea

⁶École Polytechnique Universitaire de Grenoble (Institut Universitaire de France), Université Joseph Fourier, 28 avenue Benoît Frachon, BP 53, 38041 Grenoble, France

⁷Max Planck Institute for Chemistry, Atmospheric Division, Joh.-J.-Becher-Weg 27, 55128 Mainz, Germany

⁸Unité de Formation et de Recherche de Physique et Observatoire des Sciences de l'Univers (Institut Universitaire de France), Université Joseph Fourier, Domaine Universitaire, BP 68, 38041, Grenoble, France

An iridium anomaly at the Cretaceous/Tertiary boundary layer has been attributed to an extraterrestrial body that struck the Earth some 65 million years ago¹. It has been suggested that, during this event, the carrier of iridium was probably a micrometre-sized silicate-enclosed aggregate² or the nano-phase material of the vaporized impactor³. But the fate of

platinum-group elements (such as iridium) that regularly enter the atmosphere via ablating meteoroids remains largely unknown. Here we report a record of iridium and platinum fluxes on a climatic-cycle timescale, back to 128,000 years ago, from a Greenland ice core⁴. We find that unexpectedly constant fallout of extraterrestrial matter to Greenland occurred during the Holocene, whereas a greatly enhanced input of terrestrial iridium and platinum masked the cosmic flux in the dust-laden atmosphere of the last glacial age. We suggest that nanometre-sized meteoric smoke particles^{5,6}, formed from the recondensation of ablated meteoroids in the atmosphere at altitudes >70 kilometres, are transported into the winter polar vortices by the mesospheric meridional circulation⁷ and are preferentially deposited in the polar ice caps. This implies an average global fallout of 14 ± 5 kilotons per year of meteoric smoke during the Holocene.

We determined Ir and Pt in 35 sections from the 3,028.8 m European Greenland Ice-Core Project (GRIP) ice core, drilled at Summit ($72^\circ 34' \text{ N}$, $37^\circ 37' \text{ W}$) in central Greenland⁴ (see Methods). Twenty-two sections, spanning from ~ 2 up to ~ 10 yr, were selected from the period between ~ 700 and 11,500 yr before present (BP), covering the Holocene (dating uncertainty in this period is $\sim 1\%$)⁸, and 13 sections, spanning from ~ 10 up to ~ 100 yr, were selected from the period between $\sim 13,000$ and 128,000 yr BP, covering the last glacial age (LGA) back to the last interglacial (dating uncertainty in the LGA is up to $\sim 10\%$)⁸. There is a lack of data from $\sim 4,000$ to 7,000 yr BP, owing to poor ice core quality. Low and remarkably constant concentrations are recorded in the Holocene for both Ir and Pt (averages of 0.3 and 0.6 fg g^{-1} , respectively), whereas higher and more variable values are observed during the LGA (averages of 2 and 3 fg g^{-1}). Part of the glacial–interglacial difference in concentration is caused by changes in the snow accumulation rate⁸; this rate was therefore combined with the concentration to determine the flux. The average deposition fluxes of Ir and Pt during the Holocene (8 and $15 \text{ fg cm}^{-2} \text{ yr}^{-1}$, respectively) are clearly lower by a factor of 2–3 than those during the LGA (24 and $32 \text{ fg cm}^{-2} \text{ yr}^{-1}$) (Fig. 1).

Two contrasting situations emerge from our data. In the Holocene, Ir and Pt have high crustal enrichment factors (EF_c), with mean values of 70 and 17, respectively. Here $\text{EF}_c = \{[\text{metal}]_{\text{ice}}/[\text{Al}]_{\text{ice}}\}/\{[\text{metal}]_{\text{crust}}/[\text{Al}]_{\text{crust}}\}$; $\{[\text{Ir}]_{\text{crust}}/[\text{Al}]_{\text{crust}}\} = 6.5 \times 10^{-10}$ and $\{[\text{Pt}]_{\text{crust}}/[\text{Al}]_{\text{crust}}\} = 5.2 \times 10^{-9}$ (ref. 9), and $[\text{Al}]_{\text{ice}}$ is taken from ref. 10 and references therein, and is used as crustal reference (Fig. 1). These high EF_c values show that the contribution to the fluxes of these metals from terrestrial dust is negligible during the Holocene, averaging only 2% (Ir) and 7% (Pt). Additional sources were therefore predominant during the present interglacial. Although Ir is enriched in volcanic emissions¹¹, the possibility that such a constant Ir input to Greenland ice originates from explosive volcanism can be ruled out; also, a contribution from quiescent degassing to central Greenland ice seems to be unlikely because Ir species, such as iridium hexafluoride, are thought to react rapidly with the available particulates and are deposited close to the source. In contrast, the average Ir:Pt ratio in the ice during the Holocene, 0.49 ± 0.20 , is essentially the chondritic ratio (0.49; ref. 12). This evidence, together with the high elemental correlation ($R_{\text{Ir-Pt}} = 0.67$) and the lack of correlation with Al, strongly supports a common extraterrestrial origin for Ir and Pt.

Ir and Pt can therefore be considered as equivalent proxies of extraterrestrial matter, without making necessary, as in most of the previous studies, any assumptions about their cosmic origin, and can be used to calculate the cosmic fallout to the Earth. Hence, by using the Ir or Pt chondritic concentration¹², we obtain from Greenland ice an average global extraterrestrial input during the Holocene of $78 \pm 30 \text{ kt yr}^{-1}$. The micrometeorite content measured in central Greenland¹³, based on dust of size $>0.45 \mu\text{m}$

in ice, shows that these particles constitute less than 1% of our estimate of the extraterrestrial mass accretion to the Greenland ice sheet. This lends strong support to the hypothesis that the fraction of cosmic input determined with our methodology¹⁴ in Greenland ice arises from the deposition of the nanometric meteoric smoke particles^{5,6} formed from the recondensation of ablated meteoroids in the mesosphere (see Methods). Note that if these smoke particles are less than 5 nm in radius (see below), then the number of meteoric smoke particles per gram of ice sample is more than 1×10^9 , providing excellent counting statistics.

The ablation input of $78 \pm 30 \text{ kt yr}^{-1}$ is in reasonable agreement with the meteoroid input measured with the Long Duration

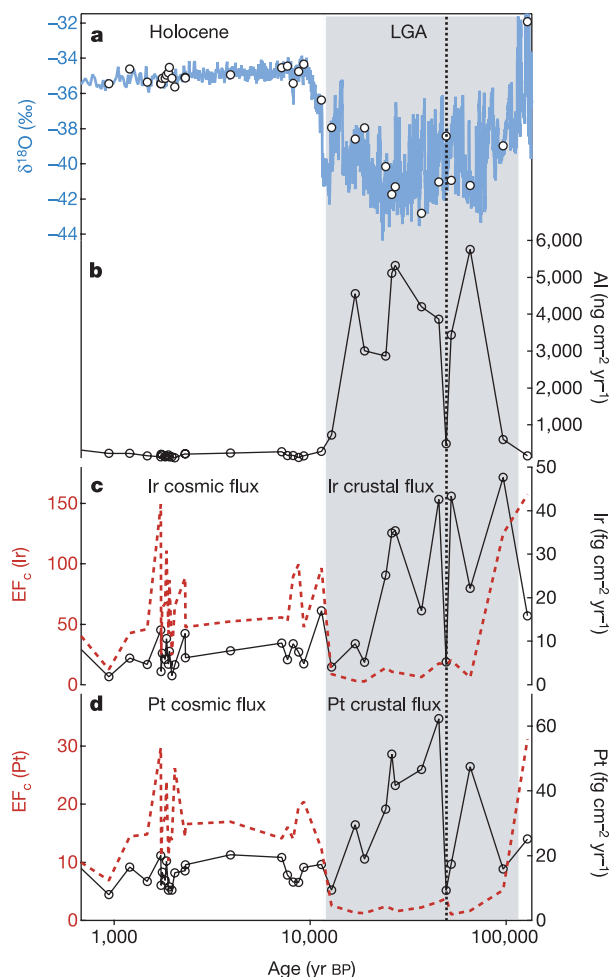


Figure 1 Change in Ir, Pt and Al depositional fluxes in central Greenland over the last climatic cycle. **a**, $\delta^{18}\text{O}$, expressed in per mil δ units ($\delta^{18}\text{O}$, light blue), used as a proxy of temperature, with less-negative values indicating higher temperatures during interglacial periods and more-negative values showing cold climatic stages. Open circles refer to values measured in our samples. **b**, Al flux (as a terrestrial dust proxy, used to calculate the crustal enrichment factor (EF_c) of Ir and Pt). **c**, Ir flux and $\text{EF}_c(\text{Ir})$ (red dotted line). **d**, Pt flux and $\text{EF}_c(\text{Pt})$ (red dotted line). Rather constant Ir and Pt fluxes are observed during the last $\sim 11,500$ yr (Holocene), while higher and more variable values are observed during the last glacial age (LGA) with a fairly synchronous variation for $\delta^{18}\text{O}$, Al, Ir and Pt (see, for instance, the reduction in metal flux and the associated $\delta^{18}\text{O}$ variation at $\sim 50,000$ yr BP, shown by the vertical dotted line). Low Ir and Pt values in the deepest sample ($\sim 128,000$ yr BP) suggest low fluxes also during the last interglacial period (Eemian). High EF_c values are observed for Ir and Pt during the Holocene, showing that contribution from crustal dust was negligible during that period and suggesting an alternative source (shown by several lines of evidence to be cosmic; see text). Conversely, much lower EF_c values are observed during the LGA, indicating that Ir and Pt were mainly derived from crustal dust during glacial times.

Table 1 Reported estimates of global extraterrestrial matter accretion rate

Method	Proxy	Type of input measured	Accretion rate (kt yr ⁻¹)
Space			
Long Duration Exposure Facility ¹⁵	Micro-impacts in space	Meteoroids	20–60
Defence satellite measurements ²⁸	Bolide detonations	Bolides	0.6*
Atmosphere			
Conventional meteor radar ¹⁹	Ablating meteoroids	Ablating meteoroids	16
Large aperture radars ¹⁷	Ablating meteoroids	Ablating meteoroids	1.6–2.7
Modelling ¹⁶	Na layer	Meteoritic smoke†	3.7–11
Glacial archives			
Collection from blue ice ²⁹	(Direct measurement in Greenland ice)	Cosmic spherules	3.2
High ice volume melting ²⁵	(Direct measurement in South Pole ice and firn)	Cosmic spherules	2.7
Instrumental neutron activation analysis ¹³	Ir in Greenland dust	Micrometeorites	0.22 (0.17–6.25)‡
ICP-SFMS (this work)	Ir and Pt in Greenland ice	Meteoritic smoke†	14§–78

* Obtained by integrating the particle flux with an infall frequency of more than once every 100 yr.

† This kind of input refers only to the ablated fraction of meteoroids, see text.

‡ Corrected for the micrometeorite under-sampling in small volumes of ice core.

§ Corrected by dividing for the latitudinal and seasonal factor, see text.

|| Integrated over the entire Earth's surface (no correction).

Exposure Facility¹⁵, an orbital impact detector placed on a spacecraft for several years, that yielded an estimate of 40 ± 20 kt yr⁻¹. This supports the current understanding that a substantial fraction of meteoroids ablates completely on entering the Earth's atmosphere¹⁶. However, our value is significantly higher than most of the recent estimates of the cosmic material influx (see Table 1 and Supplementary Fig. 2), in particular those based on incoherent scatter radar observations of meteors¹⁷ and the input required to model the global atomic Na layer in the upper mesosphere¹⁶.

An important point to consider is whether meteoric smoke is deposited uniformly over the Earth's surface, as assumed in the above estimate of the meteoric smoke input, or whether atmospheric circulation is likely to drive preferential deposition pathways. In order to examine the effects of meteoric smoke formation and atmospheric circulation on the distribution of meteoric smoke entering the troposphere, we first consider the evolution of meteoric smoke in the mesosphere. A model describing condensation, coagulation and gravitational settling (see Fig. 2 for details) shows that smoke particles do not grow large enough (>2 nm in radius) to sediment rapidly, and thus take about 45 days to descend below 60 km. However, a dominant feature of mesospheric circulation is the seasonal wind flow from the summer to the winter pole. The

velocity is typically $3\text{--}5\text{ m s}^{-1}$ between about 60 and 85 km (ref. 18), so that most of the smoke particles will be transported meridionally within the mesosphere to the winter polar vortex, before descending rapidly into the lower stratosphere. If most of the material then enters the troposphere and is deposited polewards of 55° latitude (which is about one-fifth of the global surface area), the measured input from the ice record must be divided by a factor of about five (as the meridional flow reverses every six months). This gives a global input of ablated meteoroids of 14 ± 5 kt yr⁻¹, which is in closer agreement to the estimates from conventional meteor radar measurements¹⁹ and modelling of the Na layer¹⁶ (Table 1).

The scenario is different when considering the origin of Ir and Pt during the LGA. Lower and rather constant EF_c for Ir and Pt (average values of 11 and 2, respectively; Fig. 1), and generally high elemental correlations even with Al ($R_{\text{Al-Ir}} = 0.67$; $R_{\text{Al-Pt}} = 0.87$; $R_{\text{Ir-Pt}} = 0.73$) are evidence of a noticeable contribution of Ir and Pt from the terrestrial crust during the LGA (the sample outlier for Ir:Pt of ~97,000 yr BP is excluded from this analysis). This period was characterized by a high dust loading in the Earth's atmosphere, because of enhanced dry and windy conditions and the concurrent exposure of larger areas of the continental shelf because of lower sea levels. The rather significant EF_c(Ir) during the LGA could be due to an underestimate of the [Ir]_{crust} value used for the calculation above, possibly arising from Ir spatial inhomogeneity in the Earth's crust. Alternatively, the Ir enrichment in ice could be due to the aeolian transport of Ir-enriched oxic sediments²⁰ eroded from the exposed continental shelf or to the frequent occurrence of clay minerals in the GRIP ice core during the LGA²¹. These small-sized mineral grains can be produced by the weathering of rocks relatively rich in Ir (mafic and volcanic), and are abundant in aeolian dust because of the occurrence of gravitational settling during the tropospheric transport.

The dominant fallout of terrestrial Ir and Pt into Greenland ice during the LGA means that it is difficult to test the hypothesis that there is a glacial–interglacial variation in the input of interplanetary dust, which has been proposed to account for the 100-kyr climate periodicity²². Nevertheless, when a temporarily low input of dust occurred ~50,000 yr BP during the LGA, as a result of the general instability of the climate at that time⁴ (see the Al reduction and the associated temperature variation in Fig. 1), the Ir and Pt flux returned to low holocene values. Assuming an unchanged extraterrestrial accretion rate over the Holocene and the LGA, and correcting the Ir and Pt flux for the meteoric smoke contribution, the average glacial EF_c(Pt) becomes essentially 1 and the EF_c(Ir) the more likely value of 5, therefore making clearer the dominant contribution of terrestrial Ir and Pt over the cosmic fraction. Thus, it appears that there was no significant variation in the flux of extraterrestrial matter over the LGA. Important new insights,

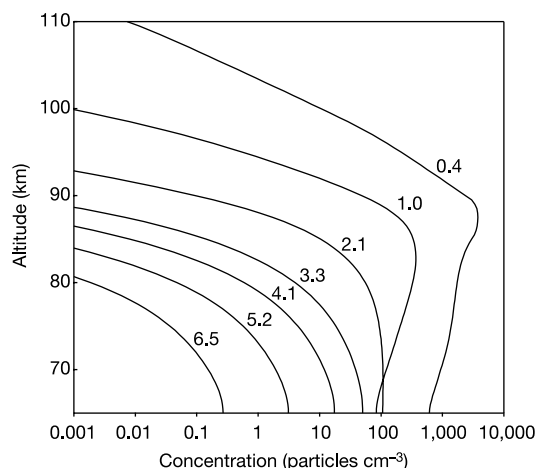


Figure 2 Vertical profiles of meteoric smoke particles of selected size. Numbers on contours show size in nm. The one-dimensional model, calculated for a mass influx of 18 kt yr^{-1} , includes meteoric ablation, polymerization of metal-containing molecules¹⁶, growth by condensation and coagulation, and sedimentation³⁰. The model shows that the meteoric smoke particles do not grow large enough (>2 nm) to sediment rapidly. It would take about 45 days for the meteoric smoke particles to sediment to below 60 km, long enough to be transported by mesospheric circulation into the polar vortices.

concerning the hypothesis of the 100-kyr cyclicity in the input of interplanetary dust, will probably be provided by measuring the concentrations of platinum group elements in Antarctic ice records such as those obtained at Dome C and Vostok, as these are relatively less affected by the input of terrestrial dust. □

Methods

Polar ice cores have provided valuable historical records of changes in heavy-metal fluxes in the Earth's atmosphere^{23,24}. Until now, the Ir occurrence in polar ice sheets was thought to be related mainly to the deposition of micrometeorites—extraterrestrial particles smaller than ~1 mm in diameter that do not ablate in the mesosphere. The few attempts to measure ultralow levels of Ir in ice cores have been restricted to particles larger than 0.45 µm, obtained from filtered samples¹³. Unfortunately, it has been shown that the typically small volumes retrieved from ice cores caused a severe under-sampling of the micrometeorite input²⁵. Statistically significant collections of micrometeorites have required large masses, of up to 8,000 t, of polar ice and firn to be investigated²⁶.

In contrast with the situation for micrometeorites, ice cores can be used to estimate the input of meteoroids that do not survive entry into the Earth's atmosphere (estimated to lie between 60% and nearly 100% of the total)¹⁶. Because of their very high entry velocities, meteoroids undergo rapid frictional heating by collision with atmospheric molecules. Their constituent minerals subsequently vaporize between an altitude of 80 and 120 km (refs 5, 6). In the middle mesosphere (50–80 km), the resulting metallic compounds are believed to polymerize with silicon oxides to form metal-rich nanoparticles known as meteoric smoke⁷.

As a consequence, platinum-group elements (PGEs), which are highly enriched in chondritic meteorites¹², are redistributed from individual submillimetric particles onto multiple nanometric particles before deposition at the Earth's surface. Thus, the concentration of PGEs in ice cores can be used to evaluate the ablated fraction of the meteoroid input. Furthermore, ice cores benefit from three relevant advantages with respect to other environmental archives for studying PGEs on a long timescale. First, the properties of PGEs remain unaltered in a low interference matrix such as ice; second, their deposition can be dated more precisely; and last, their flux can be calculated via the better known past accumulation rate.

Traditional methods for determining Ir in micrometric particulate matter filtered from ice samples¹³ fail to include the potentially significant contribution from the much smaller sized meteoric smoke. This can only be taken into account by measuring the total Ir concentration relative to the wet mass of the sample. Recently, on the basis of experience in measuring PGEs in ice cores by using inductively coupled plasma sector field mass spectrometry (ICP-SFMS)²⁷, we have developed a new method (see details in ref. 14) for quantifying the total concentration of Ir and Pt in ice, down to the sub-fg g⁻¹ level. Ultra-clean procedures and a preconcentration step have enabled us to achieve negligible procedural blanks (see also Supplementary Method 1) and procedural detection limits of 0.02 and 0.08 fg g⁻¹ for Ir and Pt, respectively¹⁴. This method benefits from an ultra-clean chemical processing of the sample based on the simple addition of ultrapure HNO₃ to the melted and pre-concentrated ice samples. The complete ionization of the meteoric smoke in the plasma (at ~7,000 K) of the ICP-SFMS is assured by the already quasi-molecular size of these particles, without the need to add any supplemental chemical reagent.

Received 20 May; accepted 18 October 2004; doi:10.1038/nature03137.

- Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. Extraterrestrial cause for the Cretaceous/Tertiary extinction. *Science* **208**, 1095–1108 (1980).
- Schuraytz, B. C. *et al.* Iridium metal in Chicxulub impact melt: forensic chemistry on the K-T smoking gun. *Science* **271**, 1573–1576 (1996).
- Wdowiak, T. J. *et al.* Presence of an iron-rich nanophase material in the upper layer of the Cretaceous-Tertiary boundary clay. *Meteorit. Planet. Sci.* **36**, 123–133 (2001).
- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice core record. *Nature* **364**, 218–220 (1993).
- Hunten, D. M., Turco, R. P. & Toon, O. B. Smoke and dust particles of meteoric origin in the mesosphere and stratosphere. *J. Atmos. Sci.* **37**, 1342–1357 (1980).
- Love, S. G. & Brownlee, D. E. Heating and thermal transformation of micrometeoroids entering the Earth's atmosphere. *Icarus* **89**, 26–43 (1991).
- Prather, M. J. & Rodriguez, J. M. Antarctic ozone: meteoric control of HNO₃. *Geophys. Res. Lett.* **15**, 1–4 (1988).
- Johnsen, S. J. *et al.* Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *Quat. Sci.* **16**, 299–307 (2001).
- Wedepohl, K. H. The composition of the continental crust. *Geochim. Cosmochim. Acta* **59**, 1217–1232 (1995).
- Hong, S., Candelone, J. P. & Boutron, C. F. Changes in zinc and cadmium concentrations in Greenland ice during the past 7760 years. *Atmos. Environ.* **31**, 2235–2242 (1997).
- Zoller, W. H., Parrington, J. R. & Kotra, J. M. P. Iridium enrichment in airborne particles from Kilauea Volcano: January 1983. *Science* **222**, 1118–1121 (1983).
- Anders, E. & Grevesse, N. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Karner, D. B. *et al.* Extraterrestrial accretion from the GISP2 ice core. *Geochim. Cosmochim. Acta* **67**, 751–763 (2003).
- Gabrielli, P. *et al.* Determination of Ir and Pt down to the sub-femtogram per gram level in polar ice by ICP-SFMS using preconcentration and a desolvation system. *J. Anal. At. Spectrom.* **19**, 831–837 (2004).
- Love, S. G. & Brownlee, D. E. A direct measurement of the terrestrial mass accretion rate of cosmic dust. *Science* **262**, 550–553 (1993).

- Plane, J. M. C. A new time-resolved model of the mesospheric Na layer: constraints on the meteor input function. *Atmos. Chem. Phys. Discuss.* **4**, 39–69 (2004).
- Mathews, J. D., Janches, D., Meisel, D. D. & Zhou, Q. H. The micrometeoroid mass flux into the upper atmosphere: Arcibco results and a comparison with prior estimates. *Geophys. Res. Lett.* **28**, 1929–1932 (2001).
- Manson, A. H. *et al.* Comparison between reference atmosphere winds and radar winds from selected locations. *Adv. Space Res.* **10**, 233–243 (1990).
- Hughes, D. W. in *Cosmic Dust* (ed. McDonnell, J. A. M.) 123–185 (Wiley, London, 1978).
- Ravizza, G. & Pyle, D. PGE and Os isotopic analysis of single sample aliquots with NIS fire assay preconcentration. *Chem. Geol.* **141**, 251–268 (1997).
- Svensson, A., Biscaye, P. E. & Grousset, F. E. Characterization of late glacial continental dust in the Greenland Ice Core Project ice core. *J. Geophys. Res.* **105**, 4637–4656 (2000).
- Muller, R. A. & MacDonald, G. J. Glacial cycles and orbital inclination. *Nature* **377**, 107–108 (1995).
- Boutron, C. F. & Patterson, C. C. Lead concentration changes in Antarctic ice during the Wisconsin/Holocene transition. *Nature* **323**, 222–225 (1986).
- Hong, S. *et al.* Climate-related variations in lead concentrations and sources in Vostok Antarctic ice from 65,000 to 240,000 years BP. *Geophys. Res. Lett.* **30**, 2138–2142 (2003).
- Peucker-Ehrenbrink, B. & Ravizza, G. The effects of sampling artifacts on cosmic dust flux estimates: A reevaluation of nonvolatile tracers (Os, Ir). *Geochim. Cosmochim. Acta* **64**, 1965–1970 (2000).
- Taylor, S., Lever, J. H. & Harvey, R. Accretion rate of cosmic spherules measured at the South Pole. *Nature* **392**, 899–903 (1998).
- Barbante, C. *et al.* Determination of Rh, Pd, and Pt in polar and alpine snow and ice by double-focusing ICPMS with microcentric nebulization. *Anal. Chem.* **71**, 4125–4133 (1999).
- Brown, P., Spalding, R. E., ReVelle, D. O., Tagliaferri, E. & Worden, S. D. The flux of small near-Earth objects colliding with the Earth. *Nature* **420**, 294–296 (2002).
- Maurette, M., Jehanno, C., Robin, E. & Hammer, C. U. Characteristics and mass distribution of extraterrestrial dust from the Greenland ice cap. *Nature* **328**, 699–702 (1987).
- Jacobson, M. Z., Lu, R., Jensen, E. J. & Toon, O. B. Modelling coagulation among particles of different composition and size. *Atmos. Environ. A* **28**, 1327–1338 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported in France by the Institut Universitaire de France, the Ministère de l'Environnement et de l'Aménagement du Territoire, the Agence de l'Environnement et de la Maîtrise de l'Energie, the Institut National des Sciences de l'Univers and the Université Joseph Fourier de Grenoble. In Italy, it was supported by the Antarctic National Research Programme, and in the UK, by the Natural Environment Research Council. We thank G. McFiggans for discussions regarding the aerosol growth model. This research was also supported by two Marie Curie fellowships of the European Community programme (IHP) and by the Alliance for Global Sustainability. This work is a contribution of GRIP organized by the European Science Foundation: we thank all the personnel working in the field in Greenland.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.B. (barbante@unive.it).

The impact of humidity above stratiform clouds on indirect aerosol climate forcing

Andrew S. Ackerman¹, Michael P. Kirkpatrick², David E. Stevens³ & Owen B. Toon⁴

¹NASA Ames Research Center, Moffett Field, California 94035, USA

²University of Tasmania, Hobart, TAS 7001, Australia

³Lawrence Livermore National Laboratory, Livermore, California 94552, USA

⁴University of Colorado, Boulder, Colorado 80309, USA

Some of the global warming from anthropogenic greenhouse gases is offset by increased reflection of solar radiation by clouds with smaller droplets that form in air polluted with aerosol particles that serve as cloud condensation nuclei¹. The resulting cooling tendency, termed the indirect aerosol forcing, is thought to be comparable in magnitude to the forcing by anthropogenic CO₂, but it is difficult to estimate because the physical processes that determine global aerosol and cloud populations are poorly understood². Smaller cloud droplets not only reflect sunlight more effectively, but also inhibit precipitation, which is expected

to result in increased cloud water^{3,4}. Such an increase in cloud water would result in even more reflective clouds, further increasing the indirect forcing. Marine boundary-layer clouds polluted by aerosol particles, however, are not generally observed to hold more water^{5–7}. Here we simulate stratocumulus clouds with a fluid dynamics model that includes detailed treatments of cloud microphysics and radiative transfer. Our simulations show that the response of cloud water to suppression of precipitation from increased droplet concentrations is determined by a competition between moistening from decreased surface precipitation and drying from increased entrainment of overlying air. Only when the overlying air is humid or droplet concentrations are very low does sufficient precipitation reach the surface to allow cloud water to increase with droplet concentrations. Otherwise, the response of cloud water to aerosol-induced suppression of precipitation is dominated by enhanced entrainment of overlying dry air. In this scenario, cloud water is reduced as droplet concentrations increase, which diminishes the indirect climate forcing.

Measuring the response of cloud water to changes in aerosol and droplet concentrations is extremely challenging because aerosol abundance typically covaries with meteorological conditions, entangling the microphysical and meteorological signals. For instance, when a regional-scale pollution plume flows offshore, cloud properties respond not only to the aerosol abundance but also to the dryness of the plume⁸.

Ship tracks, which are plumes of enhanced albedo in clouds polluted by ship exhaust, provide a natural 'laboratory' for isolating the effects of aerosol changes on cloud properties. Results from simple theoretical models^{3,4} suggest that cloud water should consistently increase in ship tracks, an expectation not generally confirmed by observations. The first reported airborne measurements of ship tracks⁹ support the theoretical expectations, with

cloud water nearly doubling as droplet concentrations (N) tripled from their background values of $\sim 30 \text{ cm}^{-3}$. Subsequent observations, however, tend to show just the opposite relationship, if any, with cloud water generally decreasing as droplet concentrations increase. For example, *in situ* measurements of over 60 ship-track interceptions during the Monterey Area Ship Tracks (MAST) field project off the coast of central California show that cloud water increases in some and decreases in others, with a slight decrease on average⁵. High-resolution, airborne remote-sensing of three ship tracks during MAST also show cloud water decreasing on average⁶. More recently, satellite observations of hundreds of ship tracks over the northeastern Pacific Ocean show a significant average decrease in cloud water⁷.

Here we investigate the dependence of cloud water on droplet concentrations using model simulations based on stratocumulus measurements during three field projects: (1) a nocturnal case study of drizzling clouds over the northeastern Atlantic during the Atlantic Stratocumulus Transition Experiment (ASTEX)^{10,11}; (2) a nocturnal study during the second Dynamics and Chemistry of Marine Stratocumulus project (DYCOMS-II) off the coast of southern California¹², with very light precipitation largely limited to the cloud deck¹³; and (3) an idealization of cloudy conditions measured over two days during the First ISCCP (International Satellite Cloud Climatology Project) Regional Experiment (FIRE-I), also off the coast of southern California. For each meteorological setting we ran a sequence of 8-h nocturnal model simulations (daytime simulations are discussed subsequently) in which only the aerosol concentration was varied. As seen in Fig. 1, only when the average precipitation rate at the surface exceeds $\sim 0.1 \text{ mm d}^{-1}$ does the average liquid-water path (LWP) increase with droplet concentrations in the simulations. This drizzling regime is consistent with simple models^{3,4} in which decreased drizzle allows LWP to increase. An increase of LWP with droplet concentrations is seen over the entire range of droplet concentrations realized for the ASTEX simulations ($\sim 25\text{--}350 \text{ cm}^{-3}$). For the DYCOMS-II and FIRE-I conditions LWP again increases with N , but only below droplet concentration thresholds of ~ 35 and 225 cm^{-3} , respectively.

These increases in LWP with N occur despite an increase with N in the rate at which the boundary layer entrains dryer air from above. The entrainment rate always increases with N in our simulations (Fig. 1), consistent with theoretical arguments that precipitation dries out cloudy air in updrafts, which reduces the moisture available for evaporative cooling of downdrafts¹⁴. Precipitation thus decreases the kinetic energy available in the boundary layer to entrain warmer air from above the temperature inversion. Conversely, a reduction in precipitation accelerates entrainment.

The response of LWP to increasing droplet concentrations can be considered to be the result of a competition between the effects of precipitation at the surface and near cloud-top. Decreased precipitation at the surface tends to increase LWP, whereas decreased precipitation near cloud-top tends to increase entrainment and thus decrease LWP. Only when the surface precipitation is sufficiently strong can it dominate the LWP response.

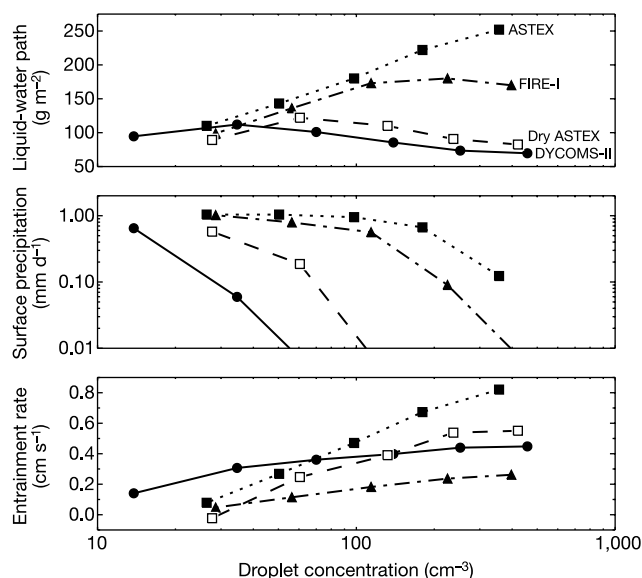


Figure 1 Liquid-water path and surface precipitation and entrainment rate as a function of cloud droplet number concentration. Liquid-water path and surface precipitation were averaged over the last two hours of 8-h simulations and the entrainment rate was averaged over eight hours. The cloud droplet number concentration was weighted by condensed water mass in each grid box. Curve labels correspond to meteorological conditions described in text. Entrainment rates ($\Delta z_i/\Delta t + Dz_i$, with z_i the horizontal average inversion height and D the large-scale horizontal wind divergence) are consistent with measured estimates of 0.38 ± 0.04 and $0.7 \pm 0.3 \text{ cm s}^{-1}$ for the DYCOMS-II and ASTEX cases^{11,12} at droplet concentrations of ~ 150 and 100 cm^{-3} .

Table 1 Meteorological conditions used for stratocumulus simulations

	ASTEX	FIRE-I	DYCOMS-II
Sea surface temperature (K)	290.4	289.0	292.5
Lifting condensation level (m)	340	250	620
Geostrophic wind speed (m s^{-1})	10	6	9
Inversion height (m)	700	600	840
Temperature increase across inversion (K)*	5.5	12	10
Moisture decrease across inversion (g kg^{-1})	1.0	3.0	7.5
Relative humidity above inversion (%)	70	40	10

*Difference in liquid-water potential temperature above and below the temperature inversion capping the boundary layer.

Precipitation depends on a number of environmental factors, and the three cases we consider here differ in a number of ways (see Table 1 and Methods). We find that the humidity of air overlying the boundary layer exerts a strong control on the surface precipitation rate. Moist air above the boundary layer is conducive to drizzle and thus favours the effect of precipitation at the surface: in the case of ASTEX the relative humidity above the boundary layer is $\sim 70\%$ and LWP monotonically increases with N over the range of droplet concentrations we consider (Fig. 1). In contrast, entrainment of dry air from above the boundary layer reduces cloud water and thereby suppresses drizzle, lowering the droplet concentration threshold above which changes in cloud water with N are dominated by the entrainment effect of precipitation. In the case of DYCOMS-II, with a relative humidity of $\sim 10\%$, LWP decreases as N increases, except at extremely low droplet concentrations. In the intermediate case of FIRE-I, with a relative humidity close to 40% , LWP is seen to increase with N only at low-to-moderate droplet concentrations, and reverses at higher droplet concentrations.

To isolate the role of humidity above the boundary layer, we ran another sequence of simulations using the ASTEX meteorology, modified with warmer, dryer air above the boundary layer. In addition to decreasing the water vapour we also increased the temperature above the boundary layer to avoid stratocumulus break-up from cloud-top entrainment instability^{15–17} (see Supplementary Fig. 1). The relative humidity above the boundary layer is $\sim 25\%$ in the modified conditions, substantially reducing precipitation and resulting in LWP decreasing as droplet concentrations increase beyond 60 cm^{-3} (Fig. 1). Thus we find that modifying the relative humidity above the boundary layer profoundly alters the balance between the competing effects of precipitation on LWP.

To clarify further the physics underlying the different responses of LWP to increasing droplet concentrations, we also ran simulations in which sedimentation of water is prevented. As seen in Fig. 2, drizzle reaches the surface for the ASTEX simulation (as observed),

providing leverage for the effect of surface precipitation on LWP. Inhibiting precipitation results in moistening of the boundary layer and increasing LWP, as found in simpler models^{3,4}. Suppressing drizzle does increase entrainment, but the entrained air is moist and does not dry the boundary layer effectively, and the cloud layer deepens as the cloud base sharpens.

For the ASTEX simulations with the same aerosol concentration but with the relative humidity reduced above the boundary layer, precipitation is drastically reduced by the entrainment of dry air. Because precipitation does not reach the surface for these conditions, it cannot be reduced further, and thus surface precipitation has no leverage on LWP. There is droplet sedimentation within the cloud layer, however; enough to restrict the entrainment rate. Completely suppressing sedimentation accelerates entrainment, drying the boundary layer and thinning the cloud layer as cloud base rises faster than cloud top.

We have also run simulations of daytime conditions, which show smaller trends in LWP as droplet concentrations increase; by offsetting longwave radiative cooling that drives stratocumulus convection, solar heating reduces convective mixing and results in a decreased response of LWP to microphysical changes (see Supplementary Table 1). In additional nocturnal simulations we find that diminished winds reduce the degree of turbulent mixing, which also results in a decreased response (Supplementary Table 1). Further simulations suggest that our results have effectively converged with respect to the grid spacing on the mesh we use (see Supplementary Fig. 2).

There are a number of implications of our findings. The response of cloud water to changes in droplet concentrations in our analysis suggests that the very concept of 'non-precipitating clouds' can be misleading. For cases in which little or no precipitation falls from a cloud layer, it is a mistake to assume that sedimentation of droplets within the cloud layer has no effect on entrainment and thus cloud water. On the contrary, it is the change in the precipitation flux from droplet sedimentation within such clouds that modulates the drying of the boundary layer by entrainment in our simulations.

Our results also show how cloud water can be depressed in clouds polluted by submicrometre aerosol particles acting as cloud nuclei, and the mechanism we describe here may well explain the decrease in LWP observed in high-resolution imagery of ship tracks⁶. (We recognize, however, that omitting partly cloudy and clear pixels may introduce a sampling bias that can lead to misleading conclusions¹⁸.) Because of the difficulties in accurately retrieving surface precipitation rates below stratocumulus using satellite measurements, we recommend that satellite retrievals of cloud microphysics (not limited to overcast pixels) be combined with meteorological analyses to assess the dependence of LWP tendencies on humidity above the boundary layer.

In contrast to one-dimensional models that include drizzle but not cloud droplet sedimentation^{3,4}, our simulations show that LWP can decrease as droplet concentrations increase. Other models that treat cloud droplet sedimentation show that when convective mixing is suppressed by strong solar heating, LWP can decrease as N increases in stratocumulus with very light precipitation^{19,20}. Those studies, however, were each limited to single meteorological settings, and did not address the role of humidity above the boundary layer. By considering a variety of meteorological settings, here we identify relative humidity above the boundary layer as a leading factor determining the response of cloud water to changes in droplet concentrations. When the air above the boundary layer is dry, we find that LWP decreases as N increases even in nocturnal simulations. Beyond marine stratocumulus, we recommend investigating the effect of overlying humidity on the response of cloud water to changes in droplet concentrations in other cloud types, such as continental and mixed-phase stratiform clouds, as well as in transitions between cloud regimes. Feedbacks with mesoscale and larger-scale circulations are also of interest.

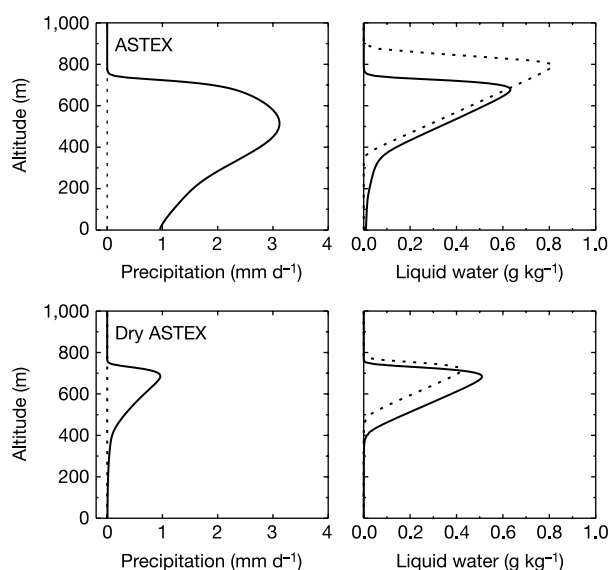


Figure 2 Horizontal average profiles of precipitation rate and liquid-water mixing ratio averaged over the last two hours of 8-h simulations. The top panels show simulations of baseline (observed) ASTEX conditions, and the bottom panels show simulations with humidity reduced above the boundary layer (see text). Dotted lines are for simulations without precipitation and solid lines for simulations with precipitation and droplet concentration of $\sim 100\text{ cm}^{-3}$ as observed¹¹. For the 'dry ASTEX' case, like the DYCOMS-II simulations and measurements¹³ at comparable droplet concentrations, precipitation is dominated by sedimentation of cloud droplets rather than larger drops produced by collisions and coalescence.

Our findings indicate that global estimates from satellite observations of the indirect aerosol forcing in boundary-layer clouds²¹ require the estimation not only of the regional changes in cloud droplet concentrations, but also of the covarying climatologies in relative humidity above stratocumulus regions. Also, predictions of climate change require global climate models to represent accurately not only such critical and challenging details as boundary-layer entrainment, but also its response to changes in cloud droplet sedimentation. Our results may help to explain why forward climate models, which consistently show an amplification of the ‘Twomey effect’²², tend to overestimate the overall indirect aerosol forcing compared to inverse calculations from simpler models constrained by the historical temperature record²³. □

Methods

Model description

The fluid dynamics model is described in ref. 24. The computational domain extends 3.4 km in both horizontal directions and 1.5 km vertically (1 km for the FIRE-I simulations), with $64 \times 64 \times 86$ cells in the x , y and z directions respectively. Grid spacing is uniform horizontally and stretched vertically to give cells of height 6 m close to the surface and in the vicinity of the inversion. The model domain is translated with the geostrophic winds, thereby reducing numerical errors associated with advection. A sponge layer occupies the upper 250 m of the domain. Subgrid-scale fluxes are modelled using a dynamic turbulence model²⁵. A modified version of the model in ref. 26 is used to treat surface-layer fluxes in the bottom 100 m of the domain. Subsidence and radiative forcings are linearly attenuated to zero in the 300 m above the inversion (defined as the average height where the total water mixing ratio exceeds a threshold that depends on the meteorological scenario) to prevent drift of the overlying atmospheric properties resulting from any unbalanced forcings. The cloud microphysics model is described in ref. 27 and the linkages with the fluid dynamics model are described elsewhere²⁸. Particle size distributions are resolved into 20 bins over a range from 0.01 to 5 μm radius for dry condensation nuclei, which are assumed to consist of ammonium bisulphate, and over a range from 1 to 500 μm radius for activated cloud droplets. The total particle number concentration at each grid point is fixed in each simulation at 40, 75, 150, 300 and 600 cm^{-3} for the ASTEX and FIRE-I progressions, and additionally at 20 cm^{-3} for the DYCOMS-II sequence. The aerosol distributions are log-normal with a geometric mean radius of 0.1 μm and a geometric standard deviation of 1.5. Radiative transfer is calculated for each column once every minute with a two-stream model²⁹.

Set-up of model simulations

The ASTEX simulations are based on an idealization of the 12–13 June 1992 nocturnal stratocumulus measurements obtained by the UK Meteorological Research Flight C-130 aircraft (flight A209) of the Atlantic Stratocumulus Transition Experiment^{10,11}, with model initialization and forcings adapted from the Fourth Global Energy and Water Experiment (GEWEX) Cloud System Studies (GCSS) Boundary Layer Cloud Workshop³⁰. We depart from the workshop specifications by initializing our model domain as initially cloud-free and using surface similarity for surface fluxes, with the sea surface temperature fixed at 292.4 K. As done in the Eighth GCSS Boundary Layer Cloud Workshop³¹ the fluxes of heat and moisture are fixed during the first two hours, in this case at 10 and 30 W m^{-2} , respectively. Instead of the radiative flux parameterization in the workshop specification we use a two-stream radiative transfer model in which the precipitable water vapour overlying the model domain is fixed at 2.2 cm, resulting in a net (upward) longwave flux of 70 W m^{-2} above the boundary layer after the cloud layer forms. For the ‘dry ASTEX’ simulations the surface moisture flux during the first two hours is increased to 60 W m^{-2} and the overlying precipitable water vapour is increased to 3.2 cm. The inversion height is defined as the average altitude where the total water mixing ratio (in grams of water per kilogram of dry air) is 8 g kg^{-1} .

The DYCOMS-II simulations are based on an idealization of nocturnal stratocumulus observations obtained during the first research flight of the second Dynamics and Chemistry of Marine Stratocumulus field study¹³, with model initialization and forcings adapted from step 4 of the Eighth GCSS Boundary Layer Cloud Workshop³¹. The precipitable water vapour overlying the model domain is fixed at 1.85 cm, resulting in a net longwave flux of 80 W m^{-2} above the boundary layer after the cloud layer forms. The total water mixing ratio defining the inversion height is 8 g kg^{-1} .

The FIRE-I simulations are based on an idealization of stratocumulus observations obtained during 14–15 July 1987 of the First ISCCP Regional Experiment with model initialization and forcings adapted from the European Project on Cloud Systems in Climate Models (EUROCS) stratocumulus case³². We use a large-scale divergence rate of $5 \times 10^{-6} \text{ s}^{-1}$ and as in other cases ignore any large-scale advective forcings. The fluxes of heat and moisture during the first two hours are fixed at 5 and 24 W m^{-2} , respectively. The precipitable water vapour overlying the model domain is fixed at 2 cm, resulting in a net longwave flux of 75 W m^{-2} above the boundary layer after the cloud layer forms. The total water mixing ratio defining the inversion height is 9 g kg^{-1} .

Received 26 May; accepted 21 October 2004; doi:10.1038/nature03174.

- Twomey, S. Pollution and the planetary albedo. *Atmos. Environ.* **8**, 1251–1256 (1974).
- Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* Ch. 5 and 6 (Cambridge Univ. Press, UK/USA, 2001).

- Albrecht, B. Aerosols, cloud microphysics, and fractional cloudiness. *Science* **245**, 1227–1230 (1989).
- Pincus, R. & Baker, M. B. Effect of precipitation on the albedo susceptibility of clouds in the marine boundary layer. *Nature* **372**, 250–252 (1994).
- Ackerman, A. S. et al. Effects of aerosols on cloud albedo: Evaluation of Twomey’s parameterization of cloud susceptibility using measurements of ship tracks. *J. Atmos. Sci.* **57**, 2684–2695 (2000).
- Platnick, S. et al. The role of background cloud microphysics in the radiative formation of ship tracks. *J. Atmos. Sci.* **57**, 2607–2624 (2000).
- Coakley, J. A. Jr & Walsh, C. D. Limits to the aerosol indirect radiative effect derived from observations of ship tracks. *J. Atmos. Sci.* **59**, 668–680 (2002).
- Brenguier, J. L. et al. Radiative properties of boundary layer clouds: Droplet effective radius versus number concentration. *J. Atmos. Sci.* **57**, 803–821 (2000).
- Radke, L. F., Coakley, J. A. Jr & King, M. D. Direct and remote sensing observations of the effects of ships on clouds. *Science* **246**, 1146–1149 (1989).
- Duykerke, P. G., Zhang, H. & Jonker, P. J. Microphysical and turbulent structure of nocturnal stratocumulus as observed during ASTEX. *J. Atmos. Sci.* **52**, 2763–2777 (1995).
- Bretherton, C. S., Austin, P. & Siems, S. T. Cloudiness and marine boundary layer dynamics in the ASTEX Lagrangian experiments: Part II: Cloudiness, drizzle, surface fluxes, and entrainment. *J. Atmos. Sci.* **52**, 2724–2735 (1995).
- Stevens, B. et al. On entrainment rates in nocturnal marine stratocumulus. *Q. J. R. Meteorol. Soc.* **129**, 3469–3493 (2003).
- Stevens, B. et al. Dynamics and chemistry of marine stratocumulus—DYCOMS-II. *Bull. Am. Meteorol. Soc.* **84**, 579–593 (2003).
- Stevens, B., Cotton, W. R., Feingold, G. & Moeng, C.-H. Large-eddy simulations of strongly precipitating, shallow, stratocumulus-topped boundary layers. *J. Atmos. Sci.* **55**, 3616–3638 (1998).
- Randall, D. A. Conditional instability of the first kind upside-down. *J. Atmos. Sci.* **37**, 125–130 (1980).
- Deardorff, J. W. Cloud top entrainment instability. *J. Atmos. Sci.* **37**, 131–147 (1980).
- MacVean, M. K. & Mason, P. J. Cloud-top entrainment instability through small-scale mixing and its parameterization in numerical models. *J. Atmos. Sci.* **47**, 1012–1030 (1990).
- Ackerman, A. S., Toon, O. B., Stevens, D. E. & Coakley, J. A. Jr. Enhancement of cloud cover and suppression of nocturnal drizzle in stratocumulus polluted by haze. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL016634 (2003).
- Ackerman, A. S., Toon, O. B. & Hobbs, P. V. Numerical modeling of ship tracks produced by injections of cloud condensation nuclei into marine stratiform clouds. *J. Geophys. Res.* **100**, 7121–7133 (1995).
- Jiang, H., Feingold, G. & Cotton, W. R. Simulations of aerosol-cloud-dynamical feedbacks resulting from entrainment of aerosol into the marine boundary layer during the Atlantic Stratocumulus Transition Experiment. *J. Geophys. Res.* **107**, doi:10.1029/2001JD001502 (2002).
- Nakajima, T., Higurashi, A., Kawamoto, K. & Penner, J. E. A possible correlation between satellite-derived cloud and aerosol microphysical parameters. *Geophys. Res. Lett.* **28**, 1171–1174 (2001).
- Haywood, J. & Boucher, O. Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review. *Rev. Geophys.* **38**, 513–543 (2000).
- Knutti, R., Stocker, T. F., Joos, F. & Plattner, G.-K. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723 (2002).
- Stevens, D. E., Ackerman, A. S. & Bretherton, C. S. Effect of domain size and numerical resolution on the simulation of shallow cumulus convection. *J. Atmos. Sci.* **59**, 3285–3301 (2002).
- Germano, M., Pionelli, U., Moin, P. & Cabot, W. H. A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids A* **3**, 1760–1765 (1991).
- Brown, A. R., Hobson, J. M. & Wood, N. Large-eddy simulation of neutral turbulent flow over rough sinusoidal ridges. *Boundary-Layer Meteorol.* **98**, 411–441 (2001).
- Ackerman, A. S., Toon, O. B. & Hobbs, P. V. A model for particle microphysics, turbulent mixing, and radiative transfer in the stratocumulus-topped marine boundary layer and comparisons with measurements. *J. Atmos. Sci.* **52**, 1204–1236 (1995).
- McFarlane, S., Evans, F. & Ackerman, A. S. A Bayesian algorithm for the retrieval of liquid water cloud properties from microwave radiometer and millimeter radar data. *J. Geophys. Res.* **107**, doi:10.1029/2001JD001011 (2002).
- Toon, O. B., McKay, C. P., Ackerman, T. P. & Santhaman, K. Rapid calculation of radiative heating and photodissociation rates in inhomogeneous multiple scattering atmospheres. *J. Geophys. Res.* **94**, 16287–16301 (1989).
- Fourth Global Energy and Water Experiment (GEWEX) Cloud System Studies (GCSS) Boundary Layer Cloud Workshop (<http://www.phys.uu.nl/~wwwimau/old/ASTEX/astexcomp.html>) (12–16 August 1996, Clermont-Ferrand, France).
- Stevens, B. et al. Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Mon. Weath. Rev.* (in the press); Eighth GCSS Boundary Layer Cloud Workshop (<http://www.atmos.ucla.edu/~bstevens/dycoms/dycoms.html>) (29–31 October 2003, Broomfield, Colorado, USA).
- Duykerke, P. G. et al. Observations and numerical simulations of the diurnal cycle of the EUROCS stratocumulus case. *Q. J. R. Meteorol. Soc.* (in the press); European Project on Cloud Systems in Climate Models (http://www.phys.uu.nl/~wwwimau/research/atm_dyn/EUROCS_PART_I/eurocs.html) (16–19 December 2002, Madrid, Spain).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Fridlind and E. Jensen for comments on the manuscript, N. Mansour, J. Ferziger, and R. Street for discussions regarding the subgrid-scale model, and B. Stevens for providing the DYCOMS-II measurements to the Eighth GCSS Boundary Layer Cloud Workshop. This work was supported by the Radiation Science Program of NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to A.S.A. (andrew.ackerman@nasa.gov).

Break-up of the Atlantic deep western boundary current into eddies at 8° S

M. Dengler, F. A. Schott, C. Eden, P. Brandt, J. Fischer & R. J. Zantopp

Leibniz-Institut für Meereswissenschaften, Düsternbrooker Weg 20, 24105 Kiel, Germany

The existence in the ocean of deep western boundary currents, which connect the high-latitude regions where deep water is formed with upwelling regions as part of the global ocean circulation, was postulated more than 40 years ago¹. These ocean currents have been found adjacent to the continental slopes of all ocean basins, and have core depths between 1,500 and 4,000 m. In the Atlantic Ocean, the deep western boundary current is estimated to carry $(10\text{--}40) \times 10^6 \text{ m}^3 \text{ s}^{-1}$ of water²⁻⁵, transporting North Atlantic Deep Water—from the overflow regions between Greenland and Scotland and from the Labrador Sea—into the South Atlantic and the Antarctic circumpolar current. Here we present direct velocity and water mass observations obtained in the period 2000 to 2003, as well as results from a numerical ocean circulation model, showing that the Atlantic deep western boundary current breaks up at 8° S. Southward of this latitude, the transport of North Atlantic Deep Water into the South Atlantic Ocean is accomplished by migrating eddies, rather than by a continuous flow. Our model simulation indicates that the deep western boundary current breaks up into eddies at the present intensity of meridional overturning circulation. For weaker overturning, continuation as a stable, laminar boundary flow seems possible.

The Atlantic deep western boundary current (DWBC) forms along the continental slope east of Greenland with its deepest layer made up by dense and cold water masses from the Denmark Strait overflow (the lower North Atlantic Deep Water, NADW). It is overlaid by overflow waters from the Iceland–Scotland sills entering the western basin through the Gibbs fracture zone in the Mid-Atlantic Ridge (the middle NADW). During its passage along the boundary of the Labrador Sea, convectively formed Labrador Sea Water joins the DWBC (the upper NADW). The DWBC carries these NADW masses southward across the Equator into the South Atlantic. The Atlantic-basin-wide meridional overturning circulation (MOC) transport is estimated to be $15\text{--}18 \text{ Sv}$ ($= 10^6 \text{ m}^3 \text{ s}^{-1}$) (refs 6, 7), considerably less than transport estimates for the DWBC at various latitudes in the Northern Hemisphere²⁻⁵. The existence of deep recirculation gyres adjacent to the western boundary is conjectured to explain such locally increased DWBC transports⁸⁻¹⁰. In the equatorial zone, the pathways of NADW are more complicated owing to a change of sign in planetary vorticity¹¹, involving large zonal excursions into the interior¹² and water mass transformations from lower into middle, and from middle into upper NADW, respectively¹³. However, the DWBC is re-established at the continental slope off Brazil at 5° S, transporting about 20 Sv of NADW southward¹⁴. Here, we analyse observations from 11° S to explore a new mechanism responsible for the southward transport of NADW at the western boundary: anticyclonic coherent eddies.

The data set used for this study involves two-year-long velocity and temperature time series from a moored array consisting of five moorings, and shipboard observations including hydrographic and direct velocity profiles. The mooring array was deployed at the Brazilian shelf break following the gradient of bathymetry to about 240 km offshore near 11° S in March 2000 (Fig. 1). The individual moorings were equipped with acoustic doppler current profilers (ADCPs) for near-surface measurements in the boundary current

regime and a variety of current meters at deeper levels (Fig. 2a). On four ship surveys, during which the moorings were serviced and redeployed, densely spaced full-depth velocity profiles and conductivity–temperature–depth–dissolved oxygen (CTD–O₂) profiles were collected by lowering an ADCP and CTD–O₂ sensors to the bottom of the ocean.

The most prominent signal in the time series from the mooring array are energetic intraseasonal fluctuations with a distinct spectral peak between 60 and 70 days. These fluctuations are very pronounced in the records from depths between 1,000 and 3,000 m, but are of smaller amplitude in the shallow and near-bottom records (Fig. 2a). The depth of the eddy kinetic energy (EKE) maximum at about 2,000 m coincides with the core of the upstream DWBC at 5° S (ref. 14). Amplitudes of the fluctuations can exceed 20 cm s^{-1} . Owing to the proximity of the measurements to the continental slope, it is convenient here to use alongshore and cross-shore velocity components that are obtained by rotating the current vectors by 36° clockwise against north. The phase relation between alongshore and cross-shore components of all energetic events is consistent with southward-translating anticyclonic eddies—that is, maximum flow away from the boundary is followed by maximum southwestward (alongshore) flow close to the boundary and then maximum shoreward flow.

As is evident from comparison with the mooring records, two of the four cruises along 11° S were carried out in March 2000 and May 2003, while southwestward flow close to the boundary was near-maximum. Sections of alongshore velocity complied from lowered ADCP profiles of these two cruises showed a pronounced dipolar structure below a depth of 1,000 m (Fig. 2b). Adjacent to the western boundary, strong southwestward flow was present with a velocity core centred at a depth of about 2,000 m, while strong northeastward flow was present farther offshore, with a velocity core at about the same depth. The maximum velocity of the offshore core was about 10 cm s^{-1} lower than its inshore counterpart. Maxima of salinity and dissolved oxygen were found within both velocity cores, indicating that more recently ventilated NADW is present in the offshore flow as well.

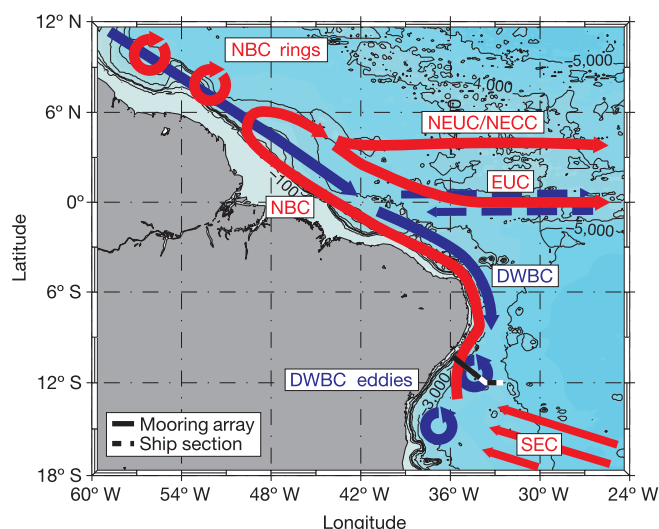


Figure 1 Circulation in the western tropical Atlantic. Schematic representation of mean currents and eddy generation at the western boundary of the tropical Atlantic with warm-water pathways in red and NADW pathways in blue. Black bar and dotted black line at 11° S indicates positions of measurement programme. Current branches indicated are the south equatorial current (SEC), the north Brazil current (NBC), the equatorial undercurrent (EUC), the north equatorial undercurrent (NEUC) merged with the north equatorial counter current (NECC) and the DWBC with alternating zonal flows marked at the Equator¹². Depth contours are also shown.

A picture consistent with both current and water-mass observations is that of eddies containing NADW, rotating anticyclonically and translating southwestward along the western boundary, thereby passing through the mooring array within 60–70 days. To quantify the individual and average structure and translation of these eddies, an eddy model was fitted to the 15 available current-meter records below 500 m depth, using subsequent 40–60 day ensembles. Eddy parameters were obtained for nine of 11 events captured by the moorings during the time span between March 2000 and February 2002. The average eddy scales and their standard deviations determined from the fit resulted in a gaussian amplitude of $51 \pm 10 \text{ cm s}^{-1}$, an e -folding radius and height of $61 \pm 10 \text{ km}$ and $1,070 \pm 200 \text{ m}$, respectively, and an eddy centre at a depth of $2,110 \pm 190 \text{ m}$. The average eddy structure obtained from the fit (Fig. 2c) is thus very similar to the velocity section measured during March 2000 (Fig. 2b). All of the individual fitted eddies rotate anticyclonically. The eddies translate southwestward with an average alongshore velocity of $3.8 \pm 0.9 \text{ cm s}^{-1}$. In addition, a shore-

ward translation of the eddies of $2.1 \pm 0.8 \text{ cm s}^{-1}$ was determined at the mooring site. This deviation from alongshore translation may be due to local and upstream topographic irregularities or to distortions of the eddies' presumed circular geometry.

For the calculation of the southward transport of deep water accomplished by the eddies, an eddy volume was defined as the volume inside the 1 cm s^{-1} isotach calculated from the individual eddy parameters excluding eddy translation velocity (Fig. 2c). On average during the two-year record, the eddies carried 25 Sv southward, of which 19 Sv is transported within the NADW layer defined here as a layer between 1,200–3,800 m. The upper and lower depth limit is comparable with the average position of the isopycnals $\sigma_1 = 32.15$ and $\sigma_4 = 45.9$ typically used to separate NADW. The NADW transport carried by the migrating eddies therefore corresponds roughly to the southward NADW transport determined at 5° S (ref. 14).

Comparing transport time series calculated from the current records and from the eddy model for a box extending from the continental slope to 170 km offshore (Fig. 2d) yields 14 Sv for the average box transport from the current records versus 15 Sv for the average eddy model transport. The agreement between both time series and their averages indicates a reasonable choice of eddy volume definition and moreover suggests that the eddies are exclusively responsible for the southward transport of NADW at the western boundary at 11° S .

Eddy-resolving general circulation models also simulate enhanced EKE at deeper levels near the western boundary at the mooring locations. In a North Atlantic version¹⁵ with about 8 km horizontal resolution of the FLAME (family of linked Atlantic model experiments) model group, a DWBC exists between 4° and 8° S , whereas south of 8° S , this DWBC breaks up into a series of anticyclonic eddies, denoted by their positive bowl-shaped temperature anomaly (Fig. 3). These eddies translate southward along the boundary with an average velocity of about 5 cm s^{-1} before they finally reach the model's open boundary at 18° S . Energy conversion terms suggest that baroclinic and barotropic instabilities are responsible for eddy generation in the DWBC region of the model. Results from a French model¹⁶ in a similar configuration, but with an extended southern model domain also show coherent eddies at the

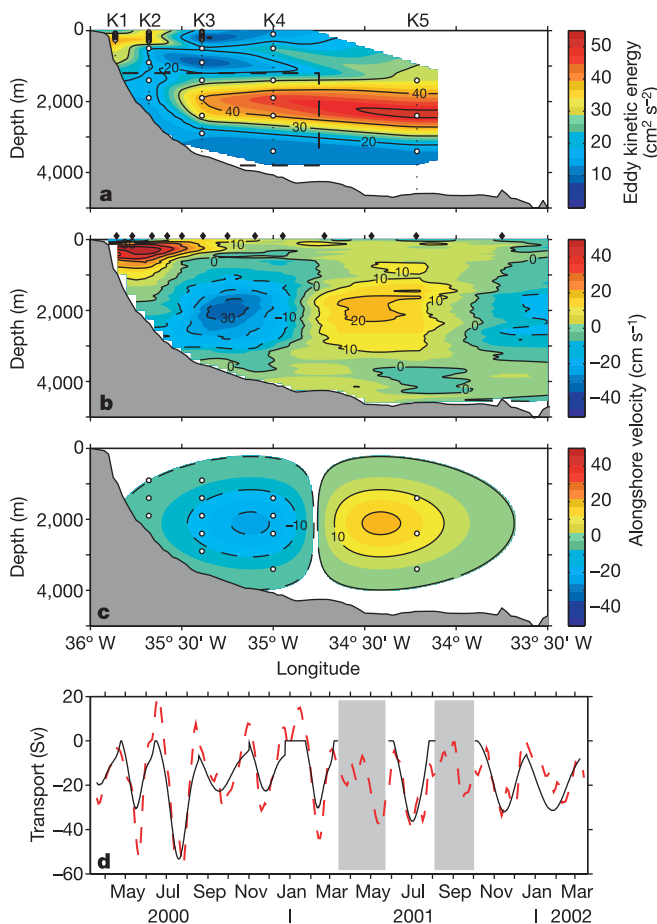


Figure 2 Distribution of eddy kinetic energy and alongshore velocity at the western boundary at 11° S . **a**, Eddy kinetic energy from the 50–90-day band-pass filtered velocity time series. Small circles indicate the distribution of moored instruments. The dashed line indicates box used for transport estimates. **b**, Snapshot of alongshore velocity measured by a lowered ADCP during March 2000; diamonds indicate positions of profile. **c**, Alongshore velocity distribution of the average eddy (the eddy-translation velocity of 3.8 cm s^{-1} is removed) obtained from fitting an eddy model to the mooring data. Small circles indicate velocity time series used for the fit. **d**, Transport time series (dashed red line) of western boundary flow for the box indicated in **a** and eddy transport calculated from the individual eddy parameters for the same box (black line). Grey shading indicates time intervals where eddy model parameters could not be quantified.

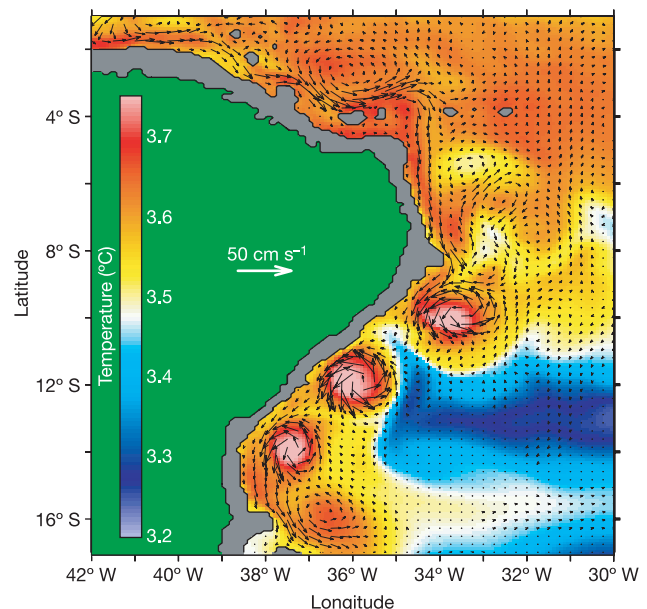


Figure 3 FLAME model results. Snapshot of velocity vectors and temperature distribution at a depth of 1,900 m during Southern Hemisphere winter from the FLAME model, using climatological forcing.

same location (A. Treguier, personal communication). Similarly, a version of the HYCOM¹⁷ with similar horizontal resolution to that of the FLAME model, but a rather different representation of the vertical coordinates, also shows eddy generation at 8° S and coherent eddies downstream (Z. Garaffo and E. Chassignet, personal communication). The FLAME model results as shown here therefore appear robust for several different model solutions.

Eddy scales and translation velocities from the FLAME model agree well with the estimates from the observations. However, the model tends to produce more vigorous eddy activity from April to September, whereas the observations do not show a similar annual cycle (Fig. 2d). The model seasonality of eddy activity is caused by seasonal variations in strength of the upstream DWBC between 4 and 8° S: during the period April–September a strong upstream DWBC leads to enhanced eddy generation, while during the period October–March a weak upstream DWBC continues as a more laminar flow south of 8° S. We speculate, on the basis of this finding, that long-term MOC variations may be related to phases of a stable, laminar flow of the DWBC followed by phases of DWBC breaking into eddies at this location during strong MOC periods.

Eddy formation in the Southern Hemisphere may be seen as a counterpart to the generation of north Brazil current (NBC) rings in the Northern Hemisphere (Fig. 1). It has been shown here that the DWBC eddies at 11° S are a major component of the cold water branch of the MOC, mirroring the role that NBC rings play for the upper limb of the Atlantic MOC^{18,19}. From observations, the fate of the DWBC and its eddies farther to the south are as yet unclear. Moored records near the boundary at several latitudes between 18° and 28° S, partially at locations of complicated topography, suggest the presence of a mean southward DWBC superimposed by large intraseasonal variability^{20,21}. At 11° S, on the other hand, an overall well-defined continental slope, the proximity to the eddy-generation region, and the combination of high-resolution shipboard measurements and moored records allowed us unambiguously to identify the existence of migrating eddies instead of a continuous DWBC, as reported here. □

Methods

Eddy fit

To quantify the individual and average structures and translation of the eddies, a multidimensional fit of a model eddy to the current-meter records was performed. The velocity structure used for the eddy description has a linear velocity gradient for its interior, as required for solid-body rotation, and an outer gaussian decay, which is consistent with the observed structures (Fig. 2b). The alongshore (V_{eddy}) and cross-shore (U_{eddy}) velocity of the eddy model were formulated as

$$V_{\text{eddy}} = v_0 + A \frac{x - (x_0 + u_0(t - t_0))}{R_0} \exp^{-\Phi} \quad (1)$$

$$U_{\text{eddy}} = u_0 + A \frac{v_0 t - v_0 t_0}{R_0} \exp^{-\Phi} \quad (2)$$

with $\Phi = \left(\frac{(x - (x_0 + u_0(t - t_0)))^2}{R_0^2} + \frac{(v_0 t - v_0 t_0)^2}{H_0^2} \right)$. Here, x is cross-shore distance, (v_0 , u_0) represents alongshore and cross-shore eddy translation, A is the amplitude of the gaussian, R_0 and H_0 are e -folding radius (the distance at which the amplitude has reduced to e^{-1}) and height and finally (x_0 , z_0) represents cross-shore distance and depth of the eddy centre. As the mooring was aligned along the cross-shore direction only, the alongshore coordinate was substituted by assuming a constant alongshore velocity for each time step $y = v_0 t$ and an alongshore eddy centre at $y_0 = v_0 t_0$. The model was fitted to subsequent 40–60-day ensembles of 40-h low-pass filtered current-meter data from below 500 m (Fig. 2c) using a multidimensional unconstrained nonlinear minimization (Nelder–Mead method). Eddy parameters could not be quantified for two time intervals during particularly weak signals (Fig. 2d).

Transport estimates

Western boundary transport was estimated from the moored velocity time series as well as from the individual eddy events that were parameterized by the fit (Fig. 2d). Data processing for transport estimates from velocity records included an empirical orthogonal function (EOF) interpolation of small temporal gaps and subsequent interpolation of the time series onto a 5 km by 20 m grid using gaussian weights. Transport was then estimated for the area 1,200 to 3,800 m deep and from the continental slope to 170 km offshore, as indicated in Fig. 2a. Finally, the transport time series was smoothed using a 10-day low-pass filter. Eddy transport was estimated by defining eddy volume as the area within the 1 cm s⁻¹ contour that was determined from the individual eddy parameters excluding eddy translation velocity. Transport was then calculated from the total velocity field (eddy

rotation and translation) inside and zero velocity outside the eddy volume. The black line in Fig. 2d represents eddy transport evaluated for the same area as used for the time series transport.

Received 3 August; accepted 19 October 2004; doi:10.1038/nature03134.

1. Stommel, H. & Arons, A. B. On the abyssal circulation of the World Ocean—I. Stationary planetary flow patterns on a sphere. *Deep-Sea Res.* **6**, 140–154 (1960).
2. Lee, T. N., Johns, W. E., Zantopp, R. J. & Fillenbaum, E. R. Moored observations of western boundary current variability and circulation at 26.5° N in the subtropical North Atlantic. *J. Phys. Oceanogr.* **26**, 962–983 (1996).
3. Schott, F. A. *et al.* Circulation and deep-water export at the western exit of the subpolar North Atlantic. *J. Phys. Oceanogr.* **34**, 817–843 (2004).
4. Johns, W. E., Fratantoni, D. M. & Zantopp, R. J. Deep western boundary current variability off northeastern Brazil. *Deep-Sea Res.* **40**, 293–310 (1993).
5. Fischer, J. & Schott, F. A. Seasonal transport variability of the Deep Western Boundary Current in the equatorial Atlantic. *J. Geophys. Res.* **102**, 27751–27769 (1997).
6. Ganachaud, A. & Wunsch, C. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* **408**, 453–457 (2000).
7. Lumpkin, R. & Speer, K. Large-scale vertical and horizontal circulation in the North Atlantic Ocean. *J. Phys. Oceanogr.* **33**, 1902–1920 (2003).
8. McCartney, M. S. Recirculation components to the deep boundary current of the northern North Atlantic. *Prog. Oceanogr.* **29**, 283–383 (1992).
9. Hogg, N. G. On the transport of the Gulf Stream between Cape Hatteras and the Grand Banks. *Deep-Sea Res.* **39**, 1231–1246 (1992).
10. Lozier, M. S. Evidence for large-scale eddy-driven gyres in the North Atlantic. *Science* **277**, 361–364 (1997).
11. Edwards, C. A. & Pedlosky, J. Dynamics of nonlinear cross-equatorial flow. Part I: Potential vorticity transformation. *J. Phys. Oceanogr.* **28**, 2382–2406 (1998).
12. Richardson, P. L. & Fratantoni, D. M. Float trajectories in the deep western boundary current and deep equatorial jets of the tropical Atlantic. *Deep-Sea Res.* **46**, 305–333 (1999).
13. Lux, M., Mercier, H. & Arhan, M. Interhemispheric exchanges of mass and heat in the Atlantic Ocean in January–March 1993. *Deep-Sea Res.* **48**, 605–638 (2001).
14. Schott, F. A., Brandt, P., Hamann, M., Fischer, J. & Stramma, L. On the boundary flow off Brazil at 5–10° S and its connection to the interior tropical Atlantic. *Geophys. Res. Lett.* **29** (17), 1840, doi:10.1029/2002GL014786 (2002).
15. Eden, C. & Böning, C. W. Sources of eddy kinetic energy in the Labrador Sea. *J. Phys. Oceanogr.* **32**, 3346–3363 (2002).
16. Treguier, A. M., Hogg, N. G., Maltrud, M., Speer, K. & Thierry, V. The origin of the deep zonal flows in the Brazil Basin. *J. Phys. Oceanogr.* **33**, 580–599 (2003).
17. Bleck, R. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Model.* **4**, 55–88 (2002).
18. Johns, W. E., Lee, T. N., Schott, F. A., Zantopp, R. J. & Evans, R. H. The North Brazil Current retroflection: Seasonal structure and eddy variability. *J. Geophys. Res.* **95**, 22103–22120 (1990).
19. Johns, W. E., Zantopp, R. J. & Goni, G. J. in *Interhemispheric Water Exchange in the Atlantic Ocean* (eds Goni, G. J. & Malanotte-Rizzoli, P.) 411–436 (Elsevier Oceanography Series, Amsterdam, 2003).
20. Weatherly, G. L., Kim, Y. Y. & Kontar, E. A. Eulerian measurements of the North Atlantic Deep Water Deep Western Boundary Current at 18° S. *J. Phys. Oceanogr.* **30**, 971–986 (2000).
21. Müller, T. J., Ikeda, Y., Zangenberg, N. & Nonato, L. V. Direct measurements of the western boundary currents off Brazil between 20° S and 28° S. *J. Geophys. Res.* **103**, 5429–5438 (1998).

Acknowledgements This study was supported by the German Bundesministerium für Bildung, Wissenschaft und Forschung (BMBF) as part of the German CLIVAR/Marin programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D. (mdengler@ifm-geomar.de).

Dating the Late Archaic occupation of the Norte Chico region in Peru

Jonathan Haas^{1*}, Winifred Creamer^{2*} & Alvaro Ruiz^{2*}

¹Department of Anthropology, The Field Museum, Chicago, Illinois 60605, USA

²Department of Anthropology, Northern Illinois University, DeKalb, Illinois 60115, USA

* These authors contributed equally to this work

The Norte Chico region on the coast of Peru north of Lima consists of four adjacent river valleys—Huaura, Supe, Pativilca and Fortaleza—in which archaeologists have been aware of a number of apparently early sites for more than 40 years (refs 1–3). To clarify the early chronology in this region, we undertook field-

work in 2002 and 2003 to determine the dates of occupation of sites in the Fortaleza and Pativilca valleys. Here we present 95 new radiocarbon dates from a sample of 13 of more than 20 large, early sites. These sites share certain basic characteristics, including large-scale monumental architecture, extensive residential architecture and a lack of ceramics. The 95 new dates confirm the emergence and development of a major cultural complex in this region during the Late Archaic period between 3000 and 1800 calibrated calendar years BC. The results help to redefine a broader understanding of the respective roles of agricultural and fishing economies in the beginnings of civilization in South America.

The Norte Chico region has greatly influenced the development of theories on the rise of complex societies in South America. Fieldwork done in the 1970s and 1980s^{4–7} at the site of Aspero at the mouth of the Supe Valley showed that this fishing community was occupied during the Late Archaic period⁸. With its several platform mound structures, Aspero inspired the theory that the initial emergence of complex society in the Andes was based on the exploitation of maritime resources rather than agriculture, a theory often referred to as the ‘maritime foundations of Andean civilization’^{3,5}. In the 1980s the first dates for a number of large inland sites in the Supe Valley suggested that these sites were occupied during the Late Archaic period⁹. Extensive excavations in the 1990s at the Supe Valley site of Caral¹⁰ firmly established the presence of a major inland, agriculturally based, component to the Late Archaic occupation in the Norte Chico^{11,12}.

Our recent work in the neighbouring Pativilca and Fortaleza has revealed that Caral and Aspero were but two of a much larger number of major Late Archaic sites in the Norte Chico. There is now evidence of an extraordinary complex of more than 20 separate major residential centres with monumental architecture concentrated in just three small valleys (Fig. 1).

To complement the earlier work in the Supe Valley and assess the overall extent of the Late Archaic occupation in the Norte Chico, we began fieldwork in the Pativilca and Fortaleza valleys in 2002 and 2003. Our limited reconnaissance in these two valleys identified at least 16 sites with surface characteristics of Late Archaic occupations: a lack of ceramics, monumental stone architecture and large circular ceremonial structures^{12,13}. These sites vary in size from ten to more than 100 hectares in area. Each has between one and seven platform mounds. The mounds, rectangular terraced pyramids, range in size from 3,000 to over 100,000 m³ (Fig. 2). Rooms were constructed on the tops and upper terraces of the structures. Another hallmark of Late Archaic sites is the sunken circular plaza. These plazas range from 20 to 40 m in diameter and are 1–2 m deep¹⁴. The sites also had large expanses of associated residential architecture, as manifested in surface indications and in stratified house floors in test pits. Excavations revealed stratified household refuse 50 to 200 cm deep. Importantly, the sites are consistently located immediately adjacent to short irrigation canals watering large tracts of land in the first terrace immediately above the river bottom.

Test excavations were conducted at a sample of 13 of the sites in the Pativilca (seven sites) and Fortaleza valleys (six sites). These excavations were designed to yield suitable material for radiocarbon dating from successive events of mound construction and from stratified levels of residential refuse. The 95 radiocarbon dates (Table 1) obtained from excavations confirmed that 11 of the 13 tested sites were occupied during the Late Archaic period (Fig. 2).

When the 95 new dates are added to the existing published dates from previous projects, there is a combined total of 127 radiocarbon dates available for the Norte Chico region, extending from 9210 to 186 calibrated calendar years BC (cal. BC). Both the earlier and later ends of this range are under-represented in this sample owing to a research focus on sites most likely to fall into the

period from 3000 to 1800 cal. BC.

Ten samples date before 3500 cal. BC. One early date of 9210 cal. BC provides limited indication of an Early Archaic occupation. The remaining nine dates confirm occupation in the region both on the coast and at inland locations between 6300 and 3500 cal. BC during the Middle Archaic period. Two of the early dates, 3710 cal. BC at Aspero⁷ and 3710 cal. BC at Porvenir are directly associated with communal architecture. The single date from Aspero is considered an anomaly⁶ and the date from Porvenir, although derived from construction material, is 700 years earlier than the next closest date. Thus it should not be inferred on the basis of these two isolated dates that mound construction had started in the fourth millennium BC.

The succeeding 700 years, between 3200 and 2500 cal. BC, with 27 dates, marks the clear appearance of large-scale communal construction and population aggregation in the Norte Chico region. Dated samples from this period have been found in direct association with the construction or occupation of platform mounds at Aspero⁶, Caral¹⁰ and Lurihuasi⁹ in the Supe Valley, at Upaca in the Pativilca Valley and at Huaricanga, Porvenir and Caballote in the Fortaleza Valley. The dates at Aspero fit well within the range of dates for the inland sites and indicate that the first construction of communal architecture occurs at both inland and coastal sites at the very beginning of the third millennium BC. On the basis of this new evidence, it is not feasible to view the maritime development at Aspero as having preceded the large-scale inland occupations. Given the significant and ongoing destruction of these early sites, including the use of Aspero as a modern landfill, it is unlikely that which site was ‘first’ can ever be known. What is clear from the new body of

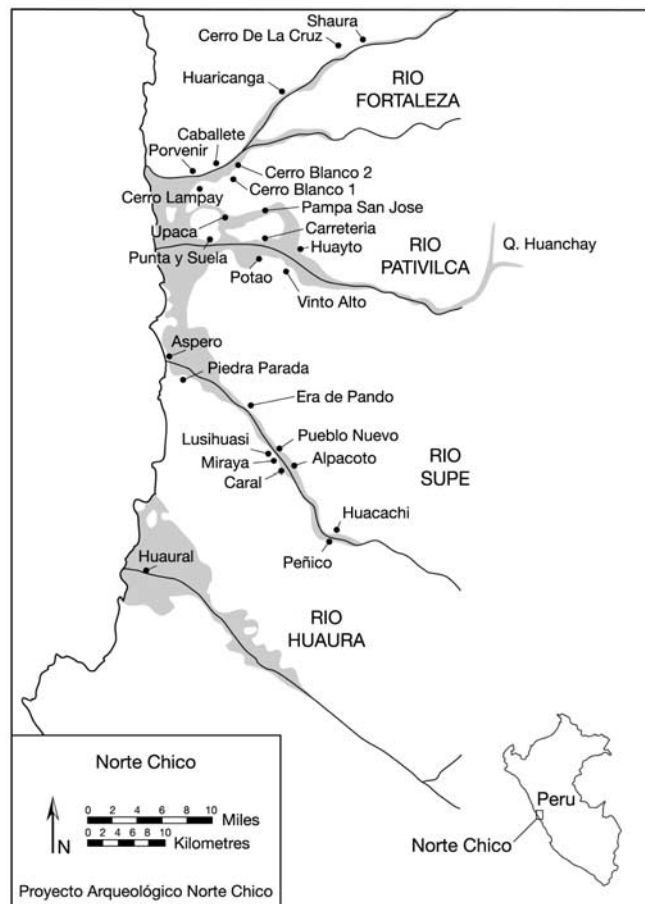


Figure 1 Map of major Late Archaic sites in the Norte Chico region.



a



b



c



d

Figure 2 Late Archaic communal architecture. **a**, Aerial photograph of Punta y Suela in the Patavilca Valley taken in 1969 (mound A in lower left is 60 m × 50 m × 12 m). **b**, Mounds A (left, 85 m × 74 m × 23 m) and B (right) at Vinto Alto in Patavilca Valley.

c, Overview of central ceremonial zone at Porvenir in Fortaleza Valley (open plaza area between mounds is 500 m across). **d**, Circle of stone stela or huancas at site of Caballote in Fortaleza Valley.

dates is that the early development of complex societies along the Peruvian coast in the Late Archaic involved an extensive inland occupation based on irrigation agriculture coupled with a more localized and much smaller-scale maritime occupation on the coast¹⁵.

The period from 2500 to 2000 cal. BC, represented by 61 dates, marks a considerable expansion of the Late Archaic occupation in the Norte Chico region. This was when most of the platform mounds were constructed, as shown by dates in this range from thirteen sites. The proliferation of sites during this period is lopsided, however, in that all the development was inland. The available dates from Aspero come only from the beginning of this period⁶. The large number and size of inland sites in the Norte Chico, all continuing to be dependent on fish and shellfish, is hard to reconcile with the scarcity of maritime sites in the same area. It therefore seems likely that residents maintained exchange relationships with maritime communities outside the Norte Chico region.

At the end of the Late Archaic and into the Initial Period, with 18 dates from 2000 to ~1500 cal. BC, some sites continued to be occupied and new sites were founded. Large multi-mound sites, typical of the preceding millennium, were abandoned and two

new site types appear: single platform complexes with attached sunken plazas, and U-shaped mound complexes. The nature and chronology of the later occupations in the Norte Chico remain to be defined.

The new dates for the Norte Chico Late Archaic provide insights into the intensity and longevity of this early development. It is now clear that Aspero, Caral and the sites in the Supe Valley were parts of a much more extensive cultural system that reached across at least three valleys and an area of 1,800 km². The concentration of monumental sites in the region is unique on the Andean landscape during the third millennium BC. The clustering of dates at the start of the third millennium BC also suggests that the Norte Chico was an important historical location, where the path of cultural evolution in the Andean region diverged from a relatively simple hunting and gathering society to a much more complex pattern of social and political organization with a mixed economy based on agriculture and marine exploitation. Domesticated plants recovered include cotton (*Gossypium barbadense*), squash (*Cucurbita* sp.), chilli (*Capsicum* sp.), beans (*Phaseolus vulgaris* and *P. lunatus*), lucuma (*Pouteria lucuma*, formerly *Lucuma obovata*), guava (*Psidium guajava*), pacay (*Inga feuillei*), camote (*Ipomoea batatas*), avocado (*Persea americana*), and achira (*Canna edulis*). While the nature of

Table 1 Radiocarbon dates from the excavations in the Pativilca and Fortaleza valleys

Site	Radiocarbon yr BP	Cal. BC	Site	Radiocarbon yr BP	Cal. BC	Site	Radiocarbon yr BP	Cal. BC	Site	Radiocarbon yr BP	Cal. BC
CAB	3,920 ± 70	2400	CB2	3,390 ± 70	1680	POR	3,780 ± 60	2210	SHA	3,660 ± 60	2030
CAB	3,590 ± 70	1940	CB2	3,630 ± 80	2000	POR	4,930 ± 70	3720	UPC	3,880 ± 60	2380
CAB	3,680 ± 70	2060	CB2	3,720 ± 90	2120	POR	3,040 ± 80	1280	UPC	4,080 ± 70	2650
CAB	3,720 ± 70	2120	CPA	3,789 ± 48	2220	POR	3,850 ± 40	2320	UPC	4,180 ± 110	2740
CAB	3,890 ± 80	2360	HCG	2,580 ± 80	670	PSJ	3,540 ± 70	1870	UPC	3,850 ± 70	2310
CAB	4,450 ± 290	3120	HCG	3,870 ± 40	2350	PSJ	3,600 ± 40	1950	UPC	3,770 ± 70	2190
CAB	3,810 ± 70	2260	HCG	3,940 ± 40	2430	PSJ	3,600 ± 60	1950	UPC	3,820 ± 70	2270
CAB	3,630 ± 70	2000	HCG	3,970 ± 110	2480	PSJ	3,710 ± 70	2100	UPC	3,820 ± 70	2270
CAB	3,670 ± 50	2050	HCG	4,110 ± 70	2690	PSJ	3,710 ± 70	2100	UPC	3,860 ± 70	2330
CAB	3,980 ± 70	2490	HCG	4,780 ± 50	3570	PSJ	3,790 ± 60	2230	UPC	2,700 ± 60	870
CAB	4,050 ± 80	2600	HCG	3,770 ± 70	2190	POT	3,215 ± 35	1480	UPC	2,910 ± 70	1110
CAB	3,740 ± 50	2140	HCG	3,860 ± 40	2330	PYS	3,775 ± 35	2200	UPC	2,910 ± 80	1110
CAB	3,330 ± 90	1620	HCG	3,910 ± 40	2390	PYS	3,935 ± 35	2420	UPC	2,160 ± 70	210
CAB	3,920 ± 70	2400	HCG	3,940 ± 70	2420	PYS	3,210 ± 70	1480	UPC	2,950 ± 70	1160
CAB	4,000 ± 70	2520	HCG	4,030 ± 70	2560	PYS	2,430 ± 70	560	VTA	3,970 ± 70	2480
CAB	4,440 ± 40	3100	HCG	4,230 ± 90	2790	PYS	2,550 ± 70	660	VTA	3,970 ± 70	2480
CAR	3,760 ± 70	2180	HCG	3,950 ± 70	2440	PYS	9,750 ± 110	9170	VTA	4,010 ± 70	2540
CB1	2,950 ± 70	1160	HYT	3,800 ± 70	2240	PYS	2,600 ± 70	750	VTA	3,700 ± 110	2100
CB1	2,960 ± 70	1180	HYT	3,820 ± 70	2270	PYS	3,520 ± 70	1840	VTA	3,860 ± 60	2330
CB1	3,080 ± 70	1330	POR	3,630 ± 70	2000	PYS	6,440 ± 70	5410	VTA	4,040 ± 70	2580
CB1	3,110 ± 70	1370	POR	3,890 ± 40	2370	PYS	6,450 ± 90	5410	VTA	3,930 ± 70	2410
CB1	3,370 ± 80	1640	POR	4,110 ± 70	2690	PYS	7,410 ± 70	6280	VTA	3,930 ± 60	2400
CB1	3,090 ± 70	1340	POR	4,160 ± 70	2740	SHA	3,540 ± 60	1870	VTA	3,940 ± 70	2420
CB1	3,420 ± 70	1720	POR	3,710 ± 70	2100	SHA	3,080 ± 70	1330			

CAB, Caballote; CAR, Carretería; CB1, Cerro Blanco 1; CB2, Cerro Blanco 2; CPA, Cemetery CP; HCG, Huaricanga; HYT, Huayto; POR, Porvenir; PSJ, Pampa San Jose; POT, Potao; PYS, Punta y Suela; SHA, Shaura; UPC, Upaca; VTA, Vinto Alto. See Methods for dating procedures.

Late Archaic society is still being defined, there are indications of social hierarchies, centralized decision-making, formalized religion, and a multifaceted economy based on inland irrigation of cotton and food plants, diverse marine resources, and a system of regular exchange of food crops, cotton, fish and shellfish^{6,9,15}.

The new dates establish that the people in this region were a significant force on the Andean cultural landscape for more than 1,200 years. By 1800 cal. BC the Norte Chico was losing its status as a focal point on the Andean landscape. Much larger polities were arising to the north and south along the coast as well as in the highlands to the east. However, basic architectural and organizational patterns first appearing in the third millennium in the Norte Chico provided a foundation for many underlying similarities that define the distinctiveness of the Andean region in the succeeding 4,000 years. □

Methodology

Survey areas and site sampling

The Pativilca Valley extends 35 km from the coast until it narrows markedly to enter the foothills of the Andes. Today, there are approximately 140 km² of land under irrigation. At the mouth of the valley it is hard to distinguish the arable land of Pativilca from that of Supe and Fortaleza, because the three share a common coastal plain. The Pativilca Valley has never been systematically surveyed, and there have been only limited, unpublished excavations at a small number of sites. The Fortaleza Valley extends 46 km from its mouth at the Pacific shoreline to the town of Chasquitambo where it narrows to enter the foothills of the Andes. Approximately 115 km² of land are under irrigated fields today.

In 1996 an archaeological survey of the middle portion of the Fortaleza Valley floor, between 20 and 1,600 m above sea level, was conducted in conjunction with construction of a power line⁹. None of the sites recorded on this survey were immediately recognized as belonging to the Late Archaic period. Reconnaissance in 2000, 2001 and 2002 revealed nine sites in the Pativilca Valley and eight sites in the Fortaleza Valley with all the hallmarks of the Late Archaic: monumental architecture, circular plazas, residential architecture and incidental surface ceramics. To obtain dates from the sites that appeared to belong to the Late Archaic and to assess their overall chronological placement, test excavations were conducted at seven of the nine in Pativilca and six of the eight sites in Fortaleza.

Two strategies were used to obtain radiocarbon samples at the sites: excavation of 1 m × 2 m test units placed in areas of stratified refuse; clearing exposed profiles left by previous construction projects and looting activities. The latter allowed for the extraction of radiocarbon samples in the interior and earlier construction phases of platform mounds and other architectural features. In all these cases, modern surfaces were completely removed to expose undisturbed, *in situ* deposits to ensure that the samples were not contaminated.

Radiocarbon dates and calibration

Dates in this paper are given in calibrated calendar dates derived from the Calib 4.4

calibration program^{16,17}. Previously published dates have been recalibrated using Calib 4.4 for consistency. Each single 'cal. BC' date is derived from the mean probability rounded to the nearest decade. All dates from excavations in the Fortaleza and Pativilca valleys are provided in the Supplementary Information, with laboratory numbers, radiocarbon years before present (that is, before 1950; yr BP), provenance and calibration ranges.

Received 17 March; accepted 26 October 2004; doi:10.1038/nature03146.

- Kosok, P. *Life, Land, and Water in Ancient Peru* 216–226 (Long Island Univ. Press, New York, 1985).
- Burger, R. *Chavín and the Origins of Andean Civilization* 27–55 (Thames and Hudson, London, 1995).
- Moseley, M. *The Incas and Their Ancestors* 2nd edn 99–160 (Thames and Hudson, London, 2001).
- Moseley, M. & Willey, G. Aspero Peru: a reexamination of the site and its implications. *Am. Antiq.* **37**, 452–468 (1973).
- Moseley, M. *The Maritime Foundations of Andean Civilization* (Cummings, Menlo Park, 1975).
- Feldman, R. *Aspero, Peru: Architecture, Subsistence Economy and Other Artifacts of a Preceramic Maritime Chiefdom*. PhD dissertation, Harvard Univ (1980).
- Feldman, R. in *Civilization in the Ancient Americas* (eds Leventhal, R. & Kolata, A.) 289–310 (Univ. New Mexico Press, Albuquerque, 1983).
- Zechenter, E. *Subsistence Strategies in the Supe Valley of the Peruvian Central Coast During the Complex Prehistoric and Initial Periods*. PhD dissertation, UCLA (1988).
- Shady, R. & Leyva, C. (eds) *La Ciudad Sagrada de Caral-Supe: Los Orígenes de la Civilización Andina y la formación del Estado Pristino en el Antiguo Perú* (Instituto Nacional de Cultura, Lima, Lima, 2003).
- Shady, R., Haas, J. & Creamer, W. Dating Caral, a preceramic urban center in the Supe Valley on the central coast of Peru. *Science* **292**, 723–726 (2001).
- Haas, J. & Creamer, W. Amplifying importance of new research in Peru, response. *Science* **294**, 1652 (2001).
- Williams, C. in *Early Ceremonial Architecture in the Andes* (ed. Donnan, C.) 227–240 (Dumbarton Oaks, Washington DC, 1985).
- Vega Centeno, R., Villacorta, L. F., Cáceres, L. E. & Marcone, G. Arquitectura monumental temprana en el valle medio de Fortaleza. *Bol. Arqueol.* **2**, 219–238 (1998).
- Williams, C. La difusión de los pozos ceremoniales en la costa peruana. *Apuntes* **1**, 1–9 (1972).
- Haas, J. & Creamer, W. in *Andean Archaeology* (ed. Silverman, H.) 35–50 (Blackwell, Oxford, 2004).
- Stuiver, M. & Reimer, P. J. Extended 14C database and revised CALIB radiocarbon calibration program. *Radiocarbon* **35**, 215–230 (1993).
- Stuiver, M. et al. INTCAL98 radiocarbon age calibration, 24000–0 cal BP. *Radiocarbon* **40**, 1041–1083 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Research for this project was supported by the National Science Foundation, the California Community Trust, The Field Museum, and the Center for Latino and Latin American Studies, Northern Illinois University. We thank the Instituto Nacional de Cultura of Lima, Peru and the municipalities of Pativilca, Paramonga and Barranca.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H. (jhaas@fieldmuseum.org).

High-quality male field crickets invest heavily in sexual display but die young

John Hunt¹, Robert Brooks¹, Michael D. Jennions², Michael J. Smith^{1*}, Caroline L. Bentsen¹ & Luc F. Bussière¹

¹School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, New South Wales, 2052, Australia

²School of Botany and Zoology, Australian National University, Canberra, ACT, 0200, Australia

* Present address: Arthur Rylah Institute for Environmental Research, 123 Brown Street, Heidelberg, Victoria, 3084, Australia

Only high-quality males can bear the costs of an extreme sexual display^{1–4}. As a consequence, such males are not only more attractive, but they often live longer than average⁵. Recent theory predicts, however, that high-quality males should sometimes invest so heavily in sexual displays that they die sooner than lower quality males^{2,6–9}. We manipulated the phenotypic quality of field crickets, *Teleogryllus commodus*, by altering the protein content of their diet. Here we show that nymphs and adult females reared on a high-protein diet lived longer than those on a low-protein diet. In contrast, adult males reared on a high-protein diet died sooner than those on low-protein diets because they invested more energy in calling during early adulthood. Our findings uphold the theoretical prediction that the relationship between longevity and sexual advertisement may be dynamic^{2,3,6–8} (that is, either positive or negative), depending on local conditions^{3,6} such as resource availability. Moreover, they caution the use of longevity as a proxy for fitness in sexual selection studies, and suggest avenues for future research on the relationship between sexual attractiveness and ageing.

The resources acquired by individual males (described as “condition” in ref. 4) may vary according to extrinsic ecological factors and genetic differences in acquisition and assimilation ability^{4,10,11}. Phenotypically plastic allocation in response to accumulated resources is thought to be the basis for condition-dependent signals of male quality^{3,4,8,11–13}. Male field crickets, for example, have been shown to decrease the time they spend calling with reduced food availability as adults^{14,15}. However, adult life-history decisions often depend critically on the resources acquired during juvenile development¹⁶.

Here, we manipulated the protein content of the diet fed to field crickets, *Teleogryllus commodus*, from hatching until death, to mimic consistent life-long differences in the ability of individuals to acquire resources. We then measured the time that males spent calling, and where possible the call structure of males, at five-day intervals. We also monitored the longevity, development time, size and weight at eclosion, as well as weight gain during the first ten days post-eclosion, of all individuals to test the prediction that males in good condition (that is, that have acquired more resources) may allocate so much energy to sexual advertisement that they die sooner than males in poorer condition.

Higher dietary protein was associated with greater nymph survival to eclosion (Fig. 1a), faster development and heavier weight at eclosion for both sexes, and with larger body size at eclosion for males (Table 1; for full details of these and all other analyses see Supplementary Information). It was also associated with greater longevity of adult females (Fig. 1b). In contrast, adult males reared on higher protein diets had shorter adult lives than those reared on a lower protein diet (Fig. 1c). This reduction in adult male longevity was associated with elevated calling effort early in life (10–20 days after eclosion) by males reared on a high-protein diet, both in terms of the proportion of males calling on a given night (Fig. 2a) and nightly calling effort (Fig. 2b). Despite dying earlier, males reared on the high-protein diet called more in their lifetime than those reared on the lower protein diets ($F_{2,198} = 47.8$, $P < 0.001$; high protein: $16,271 \pm 1,173$ s (mean $\pm$ s.e.m.); medium protein: $7,286 \pm 781$ s; low protein: $4,489 \pm 599$ s; Tukey’s post-hoc comparison: high versus medium, and high versus low, $P < 0.001$; medium versus low, $P = 0.090$). Call structure did not, however, change significantly with either diet treatment or male age and there was no significant interaction between the two (multivariate analysis of variance (MANOVA); diet treatment, Pillai’s trace = 0.096, $F_{10,128} = 0.772$, $P = 0.772$; male age, Pillai’s trace = 0.126, $F_{5,63} = 1.816$, $P = 0.122$; interaction, Pillai’s trace = 0.073, $F_{10,128} = 0.485$, $P = 0.897$). Given the effect of diet on male body size and the fact that adult size is generally correlated with call components (such as dominant frequency) in field crickets and other acoustic insects¹⁷, why our diet treatment had no effect on call structure is the subject of current investigation in our laboratory.

Three lines of evidence suggest that the higher mortality rate of males reared on high-protein diets is due to elevated calling effort in early adulthood. First, within each treatment there was a negative relationship between early calling effort (days 10–20) and longevity (Fig. 3a). Second, males on high- and medium-protein diets tended to lose weight after the fourth day post-eclosion, whereas males on the low-protein diet continued to gain or maintain their body

Table 1 Diet and sex affect development time, size and weight at eclosion and weight gain

	MANOVA					
	Pillai's trace	<i>F</i>	d.f.	<i>P</i>		
Diet	0.316	19.19	8, 818	<0.001		
Sex	0.466	89.17	4, 408	<0.001		
Diet x sex	0.098	5.29	8, 818	<0.001		
	Univariate analyses					
	Diet means			ANOVA		
	High	Medium	Low	<i>F</i> *	<i>P</i>	
Males						
Development time (days)	96 ^a (12)	98 ^a (12)	105 ^b (12)	10.64	< 0.001	
Pronotum width (mm)	6.64 ^a (0.20)	6.66 ^a (0.17)	6.52 ^b (0.23)	9.28	< 0.001	
Weight at eclosion (mg)	737 ^a (142)	722 ^{ab} (122)	674 ^b (125)	4.34	0.014	
Weight gain, days 0–10 after eclosion (mg day ^{−1})	0.5 (9.4)	2.1 (8.9)	4.6 (13.2)	2.68	0.071	
Females						
Development time (days)	97 ^a (12)	100 ^a (11)	115 ^b (17)	32.57	< 0.001	
Pronotum width (mm)	6.82 (0.23)	6.91 (0.32)	6.85 (0.27)	1.95	0.145	
Weight at eclosion (mg)	733 ^a (98)	714 ^{ab} (96)	682 ^b (99)	4.78	0.009	
Weight gain, days 0–10 after eclosion (mg day ^{−1})	17.6 ^a (15.6)	8.8 ^b (15.6)	8.1 ^b (15.6)	7.86	0.001	

The effect of the interaction between sex and diet treatment occurs because males on high- and medium-protein diets gain weight more slowly than males reared on low-protein diets, whereas the opposite is true for females. Standard deviations are provided in parentheses. Letters in superscript indicate homogeneous subsets after Tukey’s post-hoc tests. * $F_{2,200}$ (males); $F_{2,211}$ (females).

weight until day 15 (Fig. 2c), which coincided with when they started calling (Fig. 2a, b). Further support for an association between male weight loss and the onset of calling comes from the fact that in females, greater weight gain is associated with higher protein diets (Table 1). Third, males on high- and medium-protein diets lost a greater proportion of their body mass after each night's calling than males from the low-protein treatment (linear mixed model: treatment effect, $F_{2,169} = 73.18$, $P < 0.001$; least significant difference post-hoc comparisons, high versus medium, $P = 0.064$; high versus low, and medium versus low, $P < 0.001$). In the field, the negative relationship between early calling effort and longevity should, if anything, be stronger than in the laboratory because of the increased risks faced by calling males¹⁸.

Within treatments, calling effort in early adulthood (10–20 days)

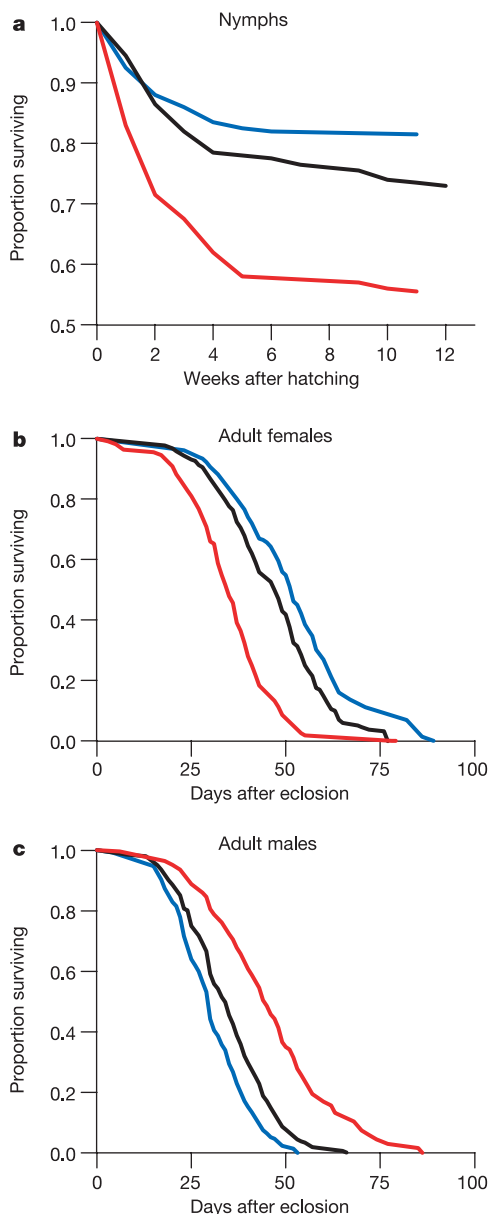


Figure 1 Nymph survival and adult longevity in the high- (blue line), medium- (black) and low-protein (red) diet treatments. **a**, Nymph survival to eclosion was highest in the high-protein treatment and lowest in the low-protein treatment (Cox regression, $\chi^2 = 29.31$, d.f. = 2, $P < 0.001$). **b**, Similarly, adult female longevity was highest in the high-protein treatment and lowest in the low-protein treatment, but **c**, the opposite was true in adult males (Cox regression, treatment $\chi^2 = 60.78$, d.f. = 2, $P < 0.001$; sex $\chi^2 = 31.06$, d.f. = 1, $P < 0.001$; treatment $\times$ sex $\chi^2 = 113.43$, d.f. = 1, $P < 0.001$).

was consistently negatively correlated with longevity (Fig. 3a). However, the relationships between late calling effort (25–30 days) and longevity were consistently positive (Fig. 3b). The shift from early to late calling across the treatments changes the relationship between mean calling effort per male per night and longevity from negative in the high- and medium-protein treatments to positive in the low-protein diet treatment (Fig. 3c). The mean condition of males in the population, therefore, not only determines how much a male invests in calling but also the nature of the

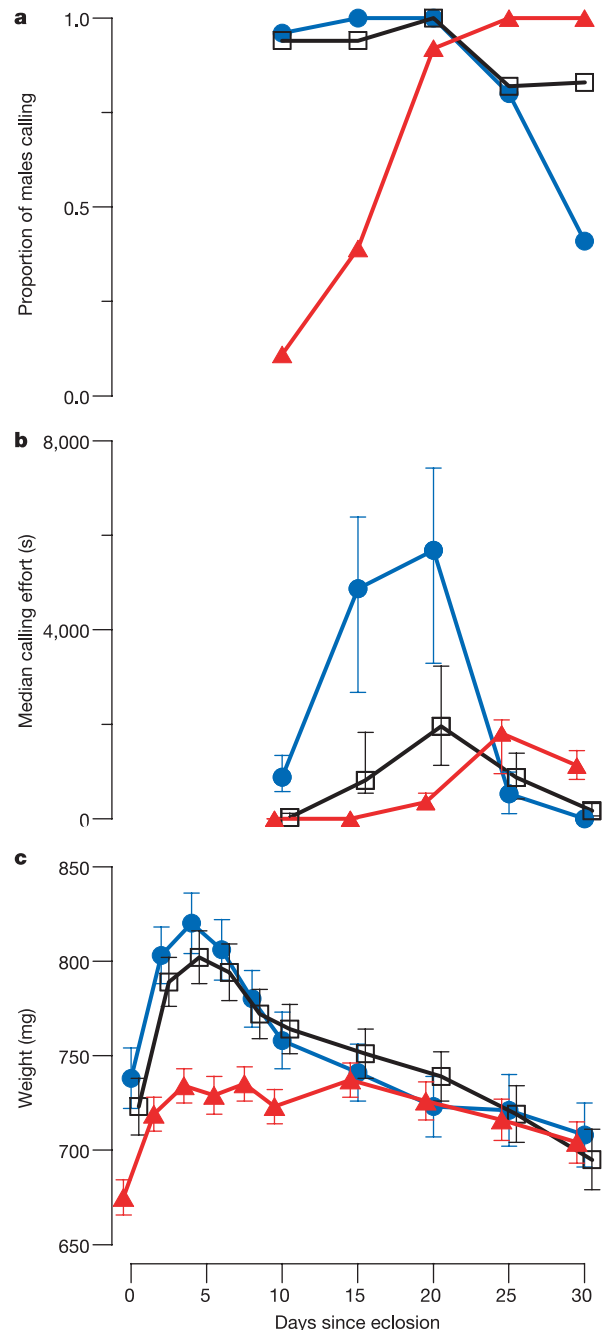


Figure 2 Reproductive effort and body weight of males in the high- (closed circles, blue line), medium- (open squares, black line) and low-protein (triangles, red line) diet treatments. **a**, Proportion of males calling. **b**, Median calling effort per night (with jack-knifed 95% confidence intervals). **c**, Male weight change ($\pm$ s.e.m.) in each treatment. There were significant differences among treatments in the quadratic and cubic coefficients describing these weight changes (treatment $\times$ (time)², $F_{2,1660} = 20.34$, $P < 0.001$; treatment $\times$ (time)³, $F_{2,1660} = 12.43$, $P < 0.001$).

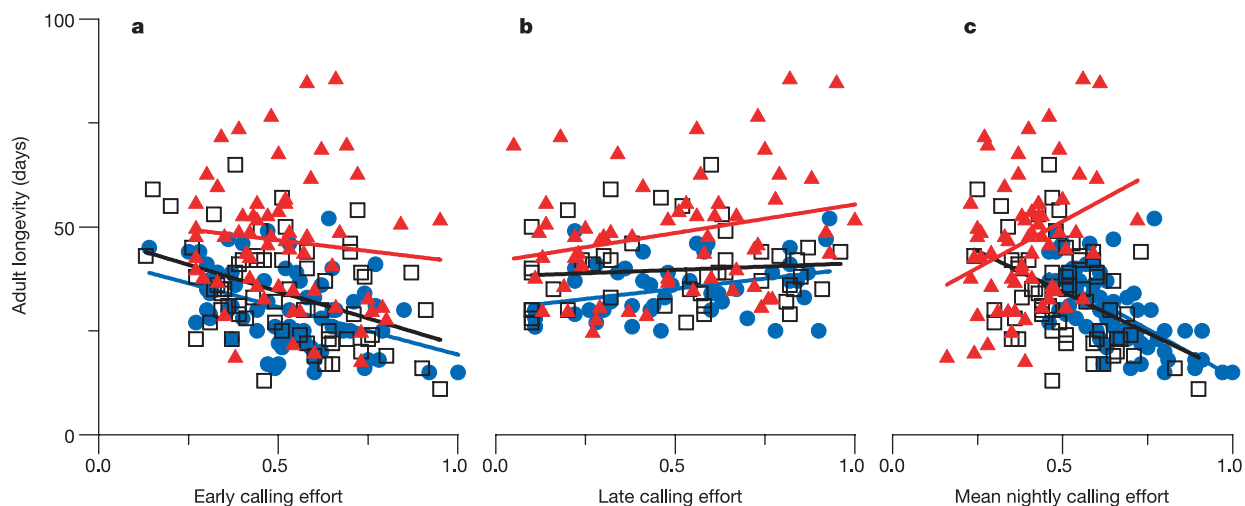


Figure 3 The relationships between longevity and early, late and mean nightly calling effort in the high- (closed circles, blue line), medium- (open squares, black line) and low-protein (triangles, red line) diet treatments (see Supplementary Information). **a**, Early calling effort: there were no treatment effects on the slope ($F_{2,192} = 0.86$, $P = 0.426$), but the common slope was significantly less than zero ($F_{1,197} = 24.38$, $P < 0.001$) and there were significant differences in elevation ($F_{2,197} = 34.09$, $P < 0.001$). **b**, Late

calling effort: there were no treatment effects on the slope ($F_{2,145} = 1.70$, $P = 0.187$), but the common slope was significantly greater than zero ($F_{1,151} = 6.62$, $P = 0.011$) and there were significant differences in elevation ($F_{2,151} = 23.06$, $P < 0.001$). **c**, Mean nightly calling effort: slopes were significantly different from one another ($F_{2,195} = 20.06$, $P < 0.001$) due to differences between the low-protein treatment and the other two treatments.

relationship between reproductive effort and longevity within populations. Our current research is examining the relative contributions of phenotypic plasticity and selection on nymph survival within treatments to the overall relationship that exists between calling effort and longevity in *T. commodus*.

Abundant indirect and limited direct evidence has shown that sexual advertisement is costly¹⁹. Nonetheless, the fundamental tenet of handicap-signalling theory that high-quality males can better bear these costs than low-quality males^{1–4}, leads to the widespread expectation that quality, signalling and longevity will be positively correlated. This viewpoint is bolstered by the fact that studies reporting a positive relationship between sexual signalling and longevity are more commonly reported in the literature⁵. For example, in the wolf spider (*Hygrolycosa rubrofasciata*) male drumming is a costly sexual display that results in weight loss, increased intrinsic mortality risk²⁰ and higher chances of predation²¹. Despite this, within treatments, faster drumming males tend to survive longer than slower drummers²⁰. This, like many similar studies⁵, has been interpreted as showing that a costly sexual display is a signal of a viability benefit to offspring²⁰. In contrast, the shorter lifespan of male crickets in our high-protein diet treatment provides the first direct confirmation of the theoretic prediction that males acquiring more resources may invest so heavily in sexual signals that they experience higher mortality than low condition males^{2,6,7}. Unlike previous studies on the costs of sexual signalling¹⁷, this relationship is not dependent on extrinsic sources of mortality (that is, predators or parasitoids¹⁶).

Dietary restriction increases lifespan in a range of organisms^{22–25}. Caloric restriction is thought to retard ageing via a number of metabolic pathways^{22,23}. The magnitude, but not the direction, of the effects of dietary restriction on longevity are often sex-specific^{26,27} and this may, in part, be due to differences in the way that the sexes allocate resources to reproduction. For example, limiting the caloric content of the diet of female *Drosophila melanogaster*^{24,27} and the protein content of the diet of female medflies (*Ceratitis capitata*)²⁵ results in reduced egg laying and concomitant increases in lifespan. However, dietary restriction also increases male longevity in these species, albeit to a lesser degree than female longevity^{22–25}, so that an overall reduction in metabolic rate can not be excluded as an alternative explanation for the

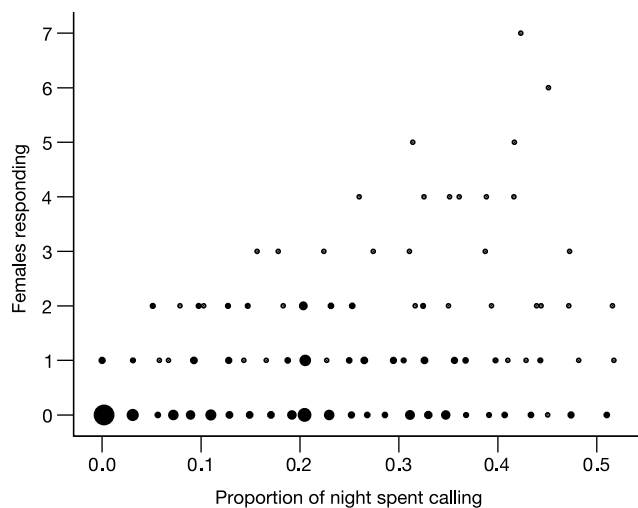


Figure 4 The number of females responding to a call in the field is positively correlated with the amount of time that the call is broadcast for. There was significant linear selection on the amount of time a call is broadcast in the field ($\beta = 0.355 \pm 0.062$ s.e., randomization test $P = 0.0001$). The largest circle represents 23 observations, the smallest represents 1 observation.

increased longevity of individuals on a restricted diet. The opposing effects of diet on male and female *T. commodus* (Fig. 1b, c) suggest that the reduced longevity of males reared on a high-protein diet is not simply a metabolic by-product, but rather the direct result of increased reproductive effort in the form of calling as a sexual advertisement early in adulthood. More importantly, females from this population impose strong positive selection on calling effort in playback experiments in the wild (Fig. 4). Collectively, these findings suggest that phenotypic plasticity in the age-specific calling effort of males is an adaptive strategy to increase mating success.

Our results demonstrate that whether the relationship between sexual advertisement and longevity is positive or negative depends critically on the mean condition of males in the population. Several authors have stressed the importance of understanding male sexual

advertisement as a form of reproductive effort within the framework of life-history theory^{6,8,10,13,28}, and of measuring such investment throughout a male's lifetime^{28–30}. Our findings provide further evidence of the value of this approach. We demonstrate that longer-lived males are not always those in the best condition and therefore our findings support theoretical claims that longevity may not always be a reliable measure of male quality^{6,9,10}. □

Methods

Laboratory feeding experiment

Animals for the laboratory feeding experiment were the F₃ descendants of 200 field-mated females collected at Smith's Lake, New South Wales, Australia, in March 2002. Cultures were provided with cat food (Friskies Go-Cat Senior) and water *ad libitum* and maintained by rearing the offspring of 100 randomly paired adults per generation in six large stock-culture containers (80 litre). Both the cultures and experimental animals were maintained in a room with a constant temperature of 28 ± 1 °C and a 10 h:14 h dark:light regime.

Experimental manipulation of condition

Manipulation of condition was achieved by feeding crickets on pellets comprising different mixtures of high-protein fish food (Pisces Enterprises, 45% protein) and oatmeal (Farmland, 12% protein). The high-, medium- and low-protein diet treatments consisted of 100% fish pellets or a dry weight mixture of 75% or 50% fish pellets with oatmeal. We created pellets by grinding fish food and oatmeal, adding water and drying the mixture in a custom-built plexiglass mould in an oven (60 °C for 12 h). Dry pellets weighed 0.121 g ± 0.002 s.e.

Experimental design

We collected 600 nymphs within 24 h of hatching and randomly assigned each to a low-, medium- or high-protein diet (200 nymphs per diet). Each cricket was housed in a separate individual plastic container (5 × 5 × 5 cm) for the duration of its life and provided with water, three diet pellets per week and a piece of egg carton for shelter. Food and water were replenished weekly and the container cleaned and nymph survival recorded. Fifth-instar nymphs were checked daily for eclosion. On the day of eclosion, and on the second, fourth, sixth, eighth and tenth day afterwards, each adult was weighed (to 0.0005 g) and pronotum width measured using an eyepiece graticule in a binocular microscope. We recorded the calling effort of living males at 10, 15, 20, 25 and 30 days after eclosion. Males were weighed the day before and the day after each calling effort measure. Food and water were replenished and the container cleaned the morning after. Females were maintained on an identical regime to males, apart from calling effort measures. Adult survival was monitored daily.

Calling effort

The calling effort of individual males was measured using an electronic monitoring device (see Supplementary Information) that monitored 64 males per night from 18:00 to 09:00 in a room set to a constant temperature of 28 ± 1 °C. Males, in their individual containers, were placed in separate styrofoam containers (15 × 10 × 10 cm), which were closed to keep males in acoustic isolation.

Field acoustic trials

We estimated the strength of sexual selection that females exert on male calling effort using playbacks of artificial calls conducted over 25 consecutive nights in the field at Smith's Lakes, New South Wales. We used SoundEdit (version 1) to construct 300 calls that varied randomly in dominant frequency, inter-call interval, number of pulses per chirp, inter-pulse interval and number of trills. We manipulated the amount of time that each call was broadcast for by randomly varying the number of times the call was repeated in the 5-min loop. Calls were broadcast from 13 sets of two speakers facing in opposite directions at 75 dB (65 cm from speakers) from 21:00 until 05:30. Females responding to a call were captured on a sticky trap (60 × 60 cm) surrounding the speakers.

Received 15 June; accepted 6 October 2004; doi:10.1038/nature03084.

1. Zahavi, A. Mate selection—a selection for a handicap. *J. Theor. Biol.* **53**, 205–214 (1975).
2. Grafen, A. Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546 (1990).
3. Nur, N. & Hasson, O. Phenotypic plasticity and the handicap principle. *J. Theor. Biol.* **110**, 275–297 (1984).
4. Rowe, L. & Houle, D. The lek paradox and the capture of genetic variance by condition dependent traits. *Proc. R. Soc. Lond. B* **263**, 1415–1421 (1996).
5. Jennions, M. D., Möller, A. P. & Petrie, M. Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* **76**, 3–36 (2001).
6. Kokko, H., Brooks, R., McNamara, J. M. & Houston, A. I. The sexual selection continuum. *Proc. R. Soc. Lond. B* **269**, 1331–1340 (2002).
7. Eshel, I., Volovik, I. & Sansone, E. On Fisher-Zahavi's handicapped sexy son. *Evol. Ecol. Res.* **2**, 509–523 (2000).
8. Höglund, J. & Sheldon, B. C. The cost of reproduction and sexual selection. *Oikos* **83**, 478–483 (1998).
9. Hansen, T. F. & Price, D. K. Good genes and old age: Do old mates provide superior genes? *J. Evol. Biol.* **8**, 759–778 (1995).
10. Hunt, J., Bussière, L. F., Jennions, M. D. & Brooks, R. What is genetic quality? *Trends Ecol. Evol.* **19**, 329–333 (2004).
11. Tomkins, J. L., Radwan, J., Kotiaho, J. S. & Tregenza, T. Genic capture and resolving the lek paradox. *Trends Ecol. Evol.* **19**, 323–328 (2004).

12. Kotiaho, J. S., Simmons, L. W. & Tomkins, J. L. Towards a resolution of the lek paradox. *Nature* **410**, 684–686 (2001).
13. Kokko, H. Good genes, old age and life-history trade-offs. *Evol. Ecol.* **12**, 739–750 (1998).
14. Wagner, W. E. J. & Hoback, W. W. Nutritional effects on male calling behaviour in the variable field cricket. *Anim. Behav.* **57**, 89–95 (1999).
15. Holzer, B., Jacot, A. & Brinkhof, M. W. G. Condition-dependent signaling affects male sexual attractiveness in field crickets, *Gryllus campestris*. *Behav. Ecol.* **14**, 353–359 (2003).
16. Roff, D. A. *Life History Evolution* (Sinauer Associates, Sunderland, Massachusetts, 2002).
17. Gerhardt, H. C. & Huber, F. *Acoustic Communication in Insects and Anurans* (Princeton Univ. Press, Princeton, 2002).
18. Zuk, M. & Kolluru, G. R. Exploitation of sexual signals by predators and parasitoids. *Q. Rev. Biol.* **73**, 415–438 (1998).
19. Kotiaho, J. S. Costs of sexual traits: a mismatch between theoretical considerations and empirical evidence. *Biol. Rev.* **76**, 365–376 (2001).
20. Kotiaho, J. S. Testing the assumptions of conditional handicap theory: costs and condition dependence of a sexually selected trait. *Behav. Ecol. Sociobiol.* **48**, 188–194 (2000).
21. Kotiaho, J., Alatalo, R. V., Mappes, J., Parri, S. & Rivero, A. Male mating success and risk of predation in a wolf spider: a balance between sexual and natural selection? *J. Anim. Ecol.* **67**, 287–291 (1998).
22. Lin, S.-J. *et al.* Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* **418**, 344–348 (2002).
23. Sohal, R. S. & Weindruch, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).
24. Chippindale, A. K., Leroi, A. M., Kim, S. B. & Rose, M. R. Phenotypic plasticity and selection in *Drosophila* life-history evolution. I. Nutrition and the cost of reproduction. *J. Evol. Biol.* **6**, 171–193 (1993).
25. Carey, J. R., Liedo, P., Müller, H.-G., Wang, J.-L. & Vaupel, J. W. Dual modes of ageing in Mediterranean fruit fly females. *Science* **281**, 996–998 (1998).
26. Müller, H.-G., Wang, J.-L., Capra, W. B., Liedo, P. & Carey, J. R. Early mortality surge in protein-deprived females causes reversal of sex differential of life expectancy in Mediterranean fruit flies. *Proc. Natl Acad. Sci. USA* **94**, 2762–2765 (1997).
27. Magwere, T., Chapman, T. & Partridge, L. Sex differences in the effect of dietary restriction on life span and mortality rates in female and male *Drosophila melanogaster*. *J. Gerontol. Biol. Sci.* **59A**, 3–9 (2004).
28. Gustafsson, L., Qvarnström, A. & Sheldon, B. C. Trade-offs between life-history traits and a secondary sexual character in male collared flycatchers. *Nature* **375**, 311–313 (1995).
29. Kokko, H. *et al.* Female choice selects for lifetime lekking performance in black grouse males. *Proc. R. Soc. Lond. B* **266**, 2109–2115 (1999).
30. Candolin, U. Changes in expression and honesty of sexual signalling over the reproductive lifetime of sticklebacks. *Proc. R. Soc. Lond. B* **267**, 2425–2430 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to S. Bertram for advice on the construction of the call-monitoring device. Thanks to J. Evans, S. Griffith, D. Gwynne, M. Head, T. Ivy, H. Kokko, J. Kotiaho and A. Lindholm for discussions and help in the field. J.H., R.B. and M.D.J. were funded by an ARC grant and L.F.B. by a NSERC Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H. (john.hunt@unsw.edu.au).

Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes

Christian Wolfrum¹, Esra Asilmaz², Edlira Luca¹, Jeffrey M. Friedman² & Markus Stoffel¹

¹Laboratory of Metabolic Diseases, and ²Laboratory of Molecular Genetics, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The regulation of fat and glucose metabolism in the liver is controlled primarily by insulin and glucagon. Changes in the circulating concentrations of these hormones signal fed or starvation states and elicit counter-regulatory responses that maintain normoglycaemia. Here we show that in normal mice, plasma insulin inhibits the forkhead transcription factor Foxa2 by nuclear exclusion and that in the fasted (low insulin) state Foxa2 activates transcriptional programmes of lipid metabolism and ketogenesis. In insulin-resistant or hyperinsulinaemic mice,

Foxa2 is inactive and permanently located in the cytoplasm of hepatocytes. In these mice, adenoviral expression of Foxa2T156A, a nuclear, constitutively active Foxa2 that cannot be inhibited by insulin¹, decreases hepatic triglyceride content, increases hepatic insulin sensitivity, reduces glucose production, normalizes plasma glucose and significantly lowers plasma insulin. These changes are associated with increased expression of genes encoding enzymes of fatty acid oxidation, ketogenesis and glycolysis. Chronic hyperinsulinaemia in insulin-resistant syndromes results in the cytoplasmic localization and inactivation of Foxa2, thereby promoting lipid accumulation and insulin resistance in the liver. Pharmacological intervention to inhibit phosphorylation of Foxa2 may be an effective treatment for type 2 diabetes.

In the postprandial state the liver takes up carbohydrates, lipids and amino acids in response to an increase in insulin, whereas in the fasted state a decrease in the plasma insulin/glucagon ratio leads to hepatic glycogen degradation, gluconeogenesis, β -oxidation and ketogenesis^{2,3}. In untreated type 1 diabetes (insulin deficiency), hyperglycaemia and ketoacidosis develop as a result of increased β -oxidation and ketogenesis⁴. In type 2 diabetes, by contrast, the hyperanabolic effects of high insulin concentrations include increased fatty acid biosynthesis and reduced β -oxidation^{5,6}. The metabolic responses to changes in plasma insulin are mediated by the insulin, phosphatidylinositol-3-OH kinase and Akt signalling pathway, which regulates the activity of several forkhead transcription factors, most notably Foxo1, an activator of rate-limiting gluconeogenic enzymes^{7,8}. The forkhead transcription factor Foxa2 is also phosphorylated in response to insulin signalling, resulting in inhibition of its transcriptional activity by nuclear exclusion¹. Here we have investigated the role of Foxa2 in response to physiological changes in plasma insulin concentration and in obesity and type 2 diabetes.

To study the role of Foxa2 in the hepatic response to changes in nutritional state, we examined the intracellular localization of Foxa2 in fasted and fed mice. In fed mice, nuclear Foxa2 protein was low and increased roughly fourfold after an overnight fast. Nuclear localization of Foxo1 was increased about twofold in the starved state (Fig. 1a). Foxa2 immunostaining was detected in hepatocyte nuclei of mice injected with PBS, whereas injection of insulin led to Foxa2 exclusion from the nucleus after 15 min (Fig. 1b). Foxa2 protein in liver nuclei of C57Bl/6J mice correlated inversely with insulin concentrations during liver perfusion studies (Fig. 1c). The regulation of nuclear and cytosolic localization of Foxo1 by insulin was less pronounced (the dissociation constant (K_d) of nuclear exclusion for Foxa2 and Foxo1 was 0.22 ± 0.03 nM and 0.74 ± 0.14 nM, respectively; Fig. 1c). We confirmed that amino acid Thr 156 in the Foxa2 protein is the site that is phosphorylated *in vivo* by immunoblotting using a specific phosphopeptide antibody. The insulin dose-dependent phosphorylation of Foxa2 correlated with Foxa2 nuclear exclusion, whereas phosphorylation of Foxo1 was observed only at 4 ng ml⁻¹ insulin (Fig. 1d).

Impairment of nuclear exclusion of Foxo1 by insulin in mouse models of insulin resistance is suggested to contribute to the overexpression of key enzymes of gluconeogenesis^{9,10}. We therefore examined the cellular localization of Foxa2 *in vivo* in three insulin-resistant mouse models: *ob/ob* mice, lipotrophic *aP2-nSrebp-1c* (*Srebp-1c*)¹¹ mice, and high-fat-induced obese C57Bl/6J (HF) mice. In contrast to normal mice, Foxa2 was localized exclusively in the cytosol of hepatocytes from both starved mice and mice fed *ad libitum* in all diabetic models (Fig. 1e), suggesting that Foxa2 is permanently inactivated in hyperinsulinaemic animals.

We considered that diverse signalling pathways might be responsible for the apparent differences in nuclear localization of Foxa2 and Foxo1. Because expression of *Irs2*, but not *Irs1*, is decreased in

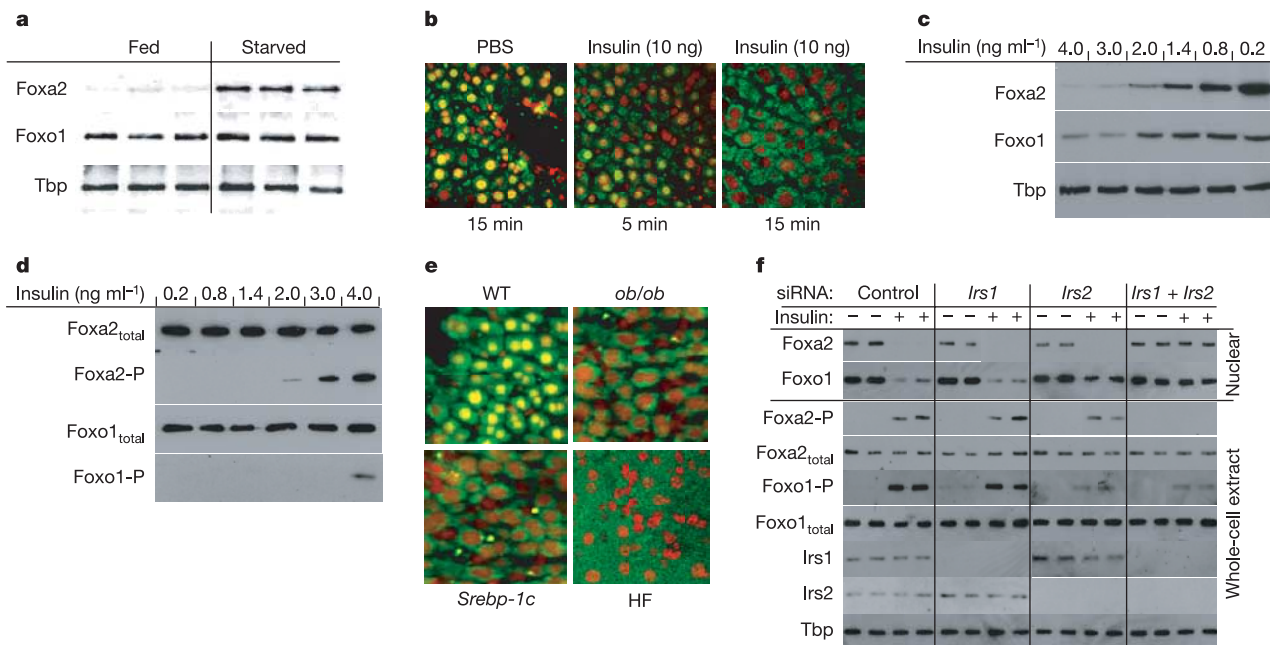


Figure 1 Insulin induces phosphorylation and nuclear exclusion of Foxa2 in mouse livers. **a**, Immunoblot analysis of Foxa2 and Foxo1 in liver nuclear extracts of fed ($n = 3$) and starved ($n = 3$) mice. The increase in nuclear expression of Foxa2 and Foxo1 in starved versus postprandial livers was 6.3 ± 0.7 ($P < 0.0001$) and 2.1 ± 0.3 ($P < 0.05$), respectively. Antibodies against TATA-binding protein (Tbp) were used as a loading control. **b**, Immunohistochemical analysis of Foxa2 expression in livers of mice perfused with either PBS or insulin (10 ng) by intraportal vein injection. Liver sections were fixed after 5 or 15 min, stained with antibodies against Foxa2 (green) and counterstained with TOPRO-3 dye (red) to visualize nuclei. **c**, Immunoblot analysis of Foxa2 and Foxo1 in nuclear extracts of mouse livers perfused with different concentrations of insulin.

d, Western blot analysis of total and phosphorylated Foxa2 and Foxo1 after immunoprecipitation and SDS-PAGE of Foxa2 and Foxo1 from whole-cell extracts of livers perfused with the indicated concentrations of insulin. **e**, Immunohistochemical assessment of livers from overnight-starved wild-type, *ob/ob*, *Srebp-1c* and high-fat-induced obese (HF) mice using antibodies against Foxa2 (green) and TOPRO-3 dye (red). **f**, Impaired nuclear exclusion and phosphorylation of Foxo1, but not Foxa2, in HepG2 cells lacking *Irs2*. Western blot analysis of nuclear and phosphorylated Foxa2 and Foxo1 in whole-cell extracts of HepG2 cells transfected with siRNAs targeting either *Irs1* or *Irs2* or both (*Irs1 + Irs2*). TATA-binding protein (Tbp) was used as a loading control.

insulin-resistant mice¹² (Supplementary Fig. 1), we tested whether Foxa2 and Foxo1 are preferentially activated through either factor. We silenced Irs1 and/or Irs2 expression in HepG2 cells by using short interfering RNAs (siRNAs) and studied nuclear localization of Foxa2 and Foxo1 in response to insulin. Insulin stimulation led to nuclear exclusion of Foxa2 in the absence of either Irs1 or Irs2, whereas Foxo1 was not translocated in cells lacking Irs2. In the absence of both Irs1 and Irs2, insulin-mediated nuclear export of both Foxa2 and Foxo1 was abolished (Fig. 1f). This indicates that the decrease in Foxo1 phosphorylation in insulin-resistant mice may be due to a defect in Irs2 signalling. By contrast, Foxa2 can be phosphorylated by either the Irs1 or the Irs2 signalling pathway.

Because mutant Foxa2T156A protein is resistant to Akt-mediated phosphorylation¹, we generated recombinant adenoviruses expressing either Foxa2 or Foxa2T156A (Ad-Foxa2 and Ad-T156A, respectively) and examined their nuclear localization in livers of wild-type and diabetic mice. Livers of fed mice injected with Ad-T156A showed nuclear staining of Foxa2 from day 1 to day 14. By contrast, Foxa2 remained in the cytoplasm of mice infected with either Ad-Foxa2 or a control adenovirus expressing green fluorescent protein (Ad-GFP; Supplementary Fig. 2a, b, and data not shown). The expression of nuclear Foxa2 in fasted and Ad-T156A-treated mice was similar (Supplementary Fig. 2c).

We examined the effect of constitutive Foxa2 activation on glucose and lipid metabolism in livers of wild-type and insulin-resistant mice. Wild-type, *ob/ob*, *Srebp-1c* and HF mice were injected with Ad-GFP, Ad-Foxa2 or Ad-T156A, and plasma glucose, insulin, triglyceride, free fatty acid and ketone body concentrations were assayed over 2 weeks and insulin tolerance was tested after 2 weeks (Fig. 2 and Supplementary Fig. 3). No significant changes in plasma glucose were observed in C57Bl/6J mice injected with either virus. However, plasma glucose concentrations were significantly decreased in diabetic mice injected with Ad-T156A (Fig. 2a). The decrease in plasma glucose was accompanied by a fall in plasma insulin (Fig. 2b) and an improvement in insulin sensi-

tivity in diabetic mice treated with Ad-T156A (Fig. 2c). Concentrations of triglycerides, free fatty acids and ketone bodies also increased during the 2-week treatment period in mice injected with Ad-T156A (Fig. 2d, e, and data not shown), whereas liver triglyceride content decreased (Fig. 2f). In addition, mice showed a significant weight loss after 14 d of treatment with Ad-T156A despite similar food and water intake. Resting O₂ consumption, CO₂ production and heat production were increased in mice expressing Foxa2T156A (data not shown).

We generated gene chip expression profiles from livers of wild-type and diabetic mice that were infected with either Ad-GFP or Ad-T156A (Supplementary Table 3). Expression profiles of key enzymes were confirmed by semi-quantitative polymerase chain reaction with reverse transcription (RT-PCR). We observed robust increases in the expression of genes involved in triglyceride degradation, fatty acid transport, mitochondrial and peroxisomal β -oxidation, ketogenesis and glycolysis (Supplementary Tables 1 and 2). The expression of glucose-6-phosphatase (G6pc), peroxisome proliferator-activated receptor- γ (Ppar- γ) and uncoupling protein-2 (Ucp-2) and Ucp-3 increased in mice infected with Ad-T156A, whereas that of Fas and Scd-1, two key regulatory enzymes involved in fatty acid synthesis, decreased. To study whether Foxa2 increased expression of these genes directly, we analysed the promoters and carried out chromatin immunoprecipitation (ChIP) and transactivation assays to confirm the functionality of the Foxa2-binding sites *in vivo* (Supplementary Fig. 4a). Foxa2 bound to Hnf3/ β sites in the promoters of genes encoding β -oxidation, ketogenesis and glycolytic enzymes, whereas Foxo1 bound only to the phosphoenolpyruvate carboxykinase (Pepck) promoter in livers of starved wild-type and *ob/ob* mice (Supplementary Fig. 4b).

To examine the physiological consequences of Foxa2-mediated gene activation, we assayed the oxidative metabolism of palmitate in liver mitochondria isolated from fed mice infected with Ad-T156A or control Ad-GFP virus. The generation of acid-soluble products (ketone bodies)¹³ was increased 2.4-fold in mitochondria of livers expressing Foxa2T156A as compared with control virus (2.95 ± 0.6 versus 1.23 ± 0.2 nmol min⁻¹ mg⁻¹, $P = 0.006$). In addition, the production of ¹⁴CO₂ from palmitate (β -oxidation) was also increased in mitochondria of livers infected with Ad-T156A (0.082 ± 0.01 versus 0.039 ± 0.01 fmol min⁻¹ mg⁻¹, $P = 0.009$).

We also examined the physiological effects of Foxa2 reactivation on hepatic glucose production and insulin signal transduction in livers of *ob/ob* mice. At high insulin concentration (20 ng ml⁻¹), there was a roughly threefold reduction in glucose output in *ob/ob* mice infected with Ad-T156A as compared with Ad-GFP (Fig. 3a). At low insulin concentration (0.5 ng ml⁻¹), hepatic glucose production was significantly increased in both groups. Hepatic Irs2 protein was significantly higher in mice injected with Ad-T156A than in those injected with Ad-GFP, and phosphorylated Akt was increased by more than twofold in mice treated with Ad-T156A when stimulated with insulin (Fig. 3b).

Our studies show that Foxa2 gain of function increases lipid metabolism and improves hepatic insulin resistance in three different strains of diabetic or insulin-resistant mice. To study the loss of function effect we characterized the phenotype of haploinsufficient Foxa2 (*Foxa2*^{+/-}) mice¹⁴. Livers of *Foxa2*^{+/-} mice fed a chow diet showed significantly decreased ketogenesis and a borderline reduction in β -oxidation. This effect was more pronounced when mice were fed a high-fat diet (Fig. 4a, b). In addition, plasma triglyceride and free fatty acid were increased in *Foxa2*^{+/-} mice as compared with wild-type littermates (data not shown). Glucose output in mice on a high-fat diet was about twofold higher in perfused livers of *Foxa2*^{+/-} as compared with wild-type littermates (Fig. 4c).

We have shown here that Foxa2 is a sensor of circulating insulin that regulates lipid metabolism and ketone body formation. In mice fed *ad libitum*, Foxa2 is mostly excluded from the nucleus. When

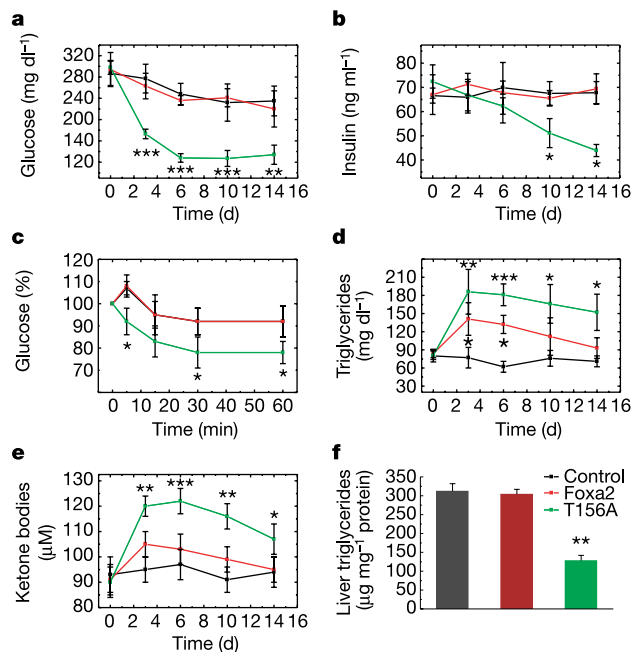


Figure 2 Metabolic measurements in *Srebp-1c* mice expressing Foxa2. *Srebp-1c* mice were infected with Ad-GFP (black), Ad-Foxa2 (red) or Ad-T156 (green) and metabolic parameters were monitored for 14 d after infection. All measurements were made after a moderate, 6-h fasting period. **a**, Plasma glucose. **b**, Plasma insulin. **c**, Percentage of starting plasma glucose concentration during an insulin tolerance test. **d**, Plasma triglyceride. **e**, Plasma ketone body. **f**, Liver triglyceride (μ g mg⁻¹ of liver protein). Values are the mean $\pm$ s.d. ($n = 5$ in each group). * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

plasma insulin decreases during fasting, Foxa2 is dephosphorylated and translocated into the nucleus, thereby inducing a transcriptional switch that activates genes involved in β -oxidation and ketogenesis. In contrast to Foxo1, the regulation of Foxa2 phosphorylation by insulin-stimulated Irs signalling is not impaired in hyperinsulinaemic or insulin-resistant mice and Foxa2 is found exclusively in the cytoplasm, despite the defect in hepatic insulin signalling in these mice. This difference can be explained, at least in part, by the preferential phosphorylation of Foxo1 through insulin-Irs2 signalling, which is impaired in hepatic insulin resistance^{9,12}, whereas Foxa2 can be phosphorylated by either insulin-Irs1 or insulin-Irs2 signalling.

Our data indicate that Foxa2 is an important regulator of β -oxidation and ketogenesis during fasting and starvation when insulin concentrations are low (glucagon, cortisol or leptin do not directly affect nuclear localization of Foxa2; data not shown), because it directly activates target genes in these pathways on nuclear translocation. The physiological significance of Foxa2 activation has been shown *in vivo* by constitutive expression of an active Foxa2 in the liver leading to increased β -oxidation, ketone

body production and increased plasma triglycerides. These findings are also supported by studies in Foxa2^{+/-} mice showing that these mice have reduced β -oxidation and ketogenesis and increased liver triglyceride content as compared with wild-type littermates when challenged with a high caloric diet. Gluconeogenesis is regulated mainly on a transcriptional level through activation of the genes encoding Pepck and G6pc by Foxo1 and its coactivator Pgc-1 (refs 7, 10, 15, 16). We found that neither Pepck gene expression nor glucose output was increased in livers of mice expressing Foxa2T156A. In addition, Foxa2, in contrast to Foxo1, did not bind to the conserved Hnf3/Foxa site in the insulin-responsive element of the Pepck promoter in ChIP assays, suggesting that Pepck is not an *in vivo* target of Foxa2.

Notably, we found increased expression of the catalytic subunits of G6pc and glucokinase (Gck), suggesting that Foxa2 may promote futile cycling between glucose-6-phosphate and glucose. The increase in expression of the rate-limiting enzymes Gck and L-type pyruvate kinase in response to Foxa2T156A is paradoxical because insulin is known to be a positive regulator of glycolysis^{17,18}. Similarly, Foxa2T156A expression activated Ucp2, although expression of this gene is known to be increased only in fed and in insulin-resistant states¹⁹. These differences are likely to reflect the existence of additional pathways that coordinately regulate other aspects of the biological response to changes in nutritional or

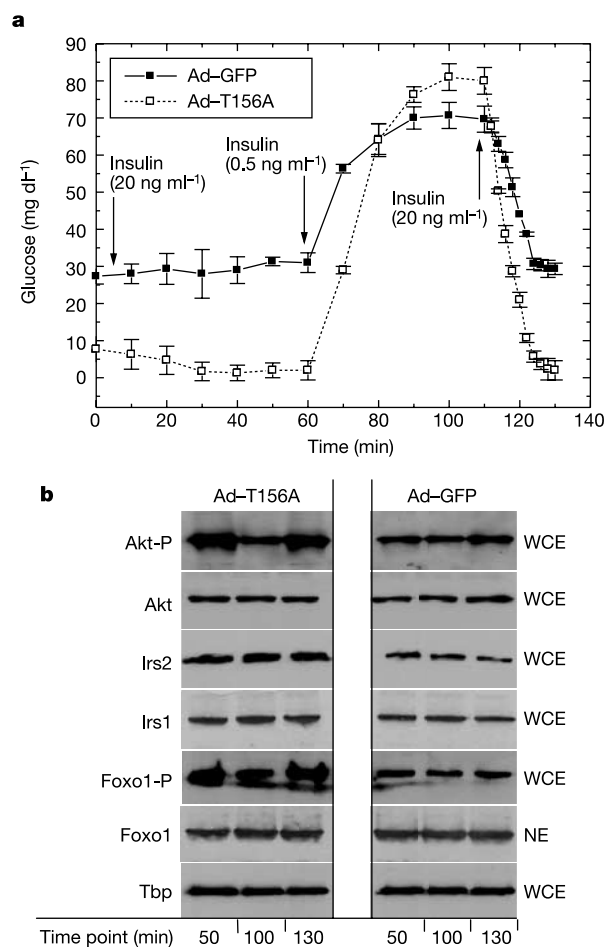


Figure 3 Improved hepatic insulin sensitivity in *ob/ob* mice infected with Ad-T156A. **a**, Glucose output from livers of *ob/ob* mice that had been infected with Ad-GFP or Ad-T156A 1 week before the study. Livers were perfused through the portal vein with a modified Krebs-Henseleit buffer and the indicated concentration of insulin. Glucose concentrations in the effluent were assayed by the glucose oxidase method. Values are the mean $\pm$ s.d. ($n = 3$). **b**, Analysis of phosphorylated Akt, total Akt and Irs2 expression by western blot. Whole-cell liver extracts (WCE) or nuclear extracts (NE) were prepared at the 50-, 100- and 130-min time points from perfused livers of mice infected with Ad-GFP or Ad-T156A (see Fig. 3a). Protein (20 μ g) was separated by SDS-PAGE and transferred to a nitrocellulose membrane. Membranes were probed with antibodies against phosphorylated Akt and Foxo1, total Akt, Irs1, Irs2 and TATA-binding protein (Tbp) as a loading control.

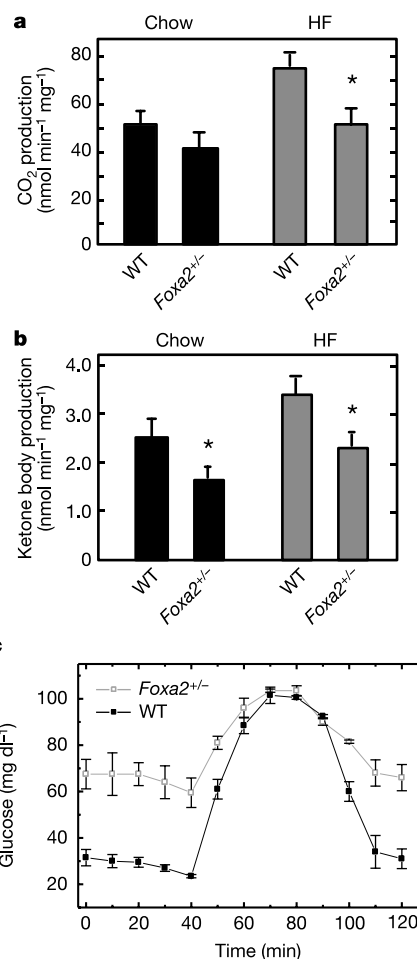


Figure 4 Decreased β -oxidation and ketogenesis and reduced hepatic insulin sensitivity in Foxa2^{+/-} mice. **a**, **b**, Production of ¹⁴CO₂ as a measure for β -oxidation (**a**) and ketone body generation (**b**) from [1-¹⁴C]palmitic acid by mitochondria from livers of wild-type (WT) or Foxa2^{+/-} mice fed a normal (chow) or high-fat (HF) diet. **c**, Glucose output measurements from livers of Foxa2^{+/-} or wild-type mice fed a high-fat diet. Livers were perfused with buffer containing high (20 ng ml⁻¹) or low (0.5 ng ml⁻¹) concentrations of insulin. Values are the mean $\pm$ s.d. ($n = 3$).

hormonal state and suggest that Foxa2 is not the principal activator of these genes in the postprandial state, thereby underlining the complexity of transcriptional regulation of glycolytic genes by positive (insulin and glucose, among others) and negative (for example, glucagon) signals.

Foxa2T156A mediates a powerful anti-diabetic response by improving glucose and lipid homeostasis in the liver of obese and diabetic mouse models. This is due to the pleiotrophic actions of Foxa2 on the expression of genes regulating lipid metabolism and glycolysis. The increase in insulin sensitivity in mice treated with Ad-T156A may be influenced by the activation of genes involved in the breakdown of liver triglycerides, fatty acid oxidation and uncoupling, as well as by a reduction in expression of key genes involved in fatty acid synthesis (e.g. *Fas* and *Scd-1*), thus improving hepatic steatosis, an outcome that has been shown to improve hepatic insulin sensitivity in different mouse models^{20,21}. Other mechanisms, such as the increased expression of Hnf4 α and Ppar- γ or increased phosphorylation of Foxo1 in mice treated with Ad-T156A, may also contribute to enhanced insulin sensitivity through the effect of these proteins on glucose and lipid metabolism^{22–24}.

Our study may have implications for the aetiology and treatment of individuals with diabetes. Hyperinsulinaemia is a common feature of obesity and insulin resistance and may elicit a vicious cycle by excluding Foxa2 from the nucleus and preventing optimal stimulation of triglyceride degradation and fatty acid oxidation. The prevention of phosphorylation of Foxa2 by pharmacological approaches may be an attractive therapeutic target for the treatment of obesity and diabetes. □

Methods

Mouse and metabolic studies

All mouse models were maintained in the C57Bl/6J background and kept on a 12-h/12-h light/dark cycle in a pathogen-free animal facility. Groups of mice were fed a high-fat diet (Harland Teklad) containing 50% fat for 6 or 12 weeks. Oxygen consumption, CO₂ and heat production, and food and water intake were simultaneously determined for four mice per experiment in an Oxymax metabolic chamber system (Columbus Instruments).

Adenovirus generation

We generated adenoviruses by using the Rapid Adenovirus Production System (Viraguest). Viruses were designed to express GFP from an independent promoter in addition to Foxa2 or Foxa2T156A (Ad-Foxa2 and Ad-Foxa2T156A, respectively). For *in vivo* experiments, mice were injected with 1×10^{11} particles of adenovirus. Virus expressing only GFP served as control (Ad-GFP).

Generation of anti-phosphopeptide antibodies

Polyclonal antibodies were produced by immunizing rabbits with a synthetic phosphorylated peptide (coupled to KLH) corresponding to the residues surrounding Thr 156 of human Foxa2. Antibodies were purified by protein A and peptide affinity chromatography (Cell Signaling).

Electrophoretic mobility shift and ChIP assays

Nuclear extracts from livers (20 μ g), prepared as described²⁵, were incubated with ³²P-labelled double-stranded oligonucleotide probes homologous to either the Foxa-binding site from the *IGFBP* promoter²⁶ or the HNF1-binding site from the *SHP* promoter²⁷. Supershift analysis was done by incubating the nuclear extracts with antibodies to Foxa2 or to HNF1 α (Cell Signaling). We carried out ChIP analysis on isolated primary hepatocytes from C57Bl/6J and *ob/ob* mice infected with either Ad-GFP or Ad-T156A by using a ChIP Assay kit (Cell Signaling) in accordance with the manufacturer's protocol. Foxa2 and Foxo1 (Cell Signaling) were precipitated with the respective antibodies.

Transfection and transactivation assays

Transfection of HepG2 cells and luciferase transactivation assays were done as described¹.

RNA silencing

HepG2 cells were grown to 60–70% confluence and transfected with siRNAs for *Irs1* (5'-NNAAGAGGUCUGGCAAGUGAU-3') and/or *Irs2* (5'-NNACAACAACAACAACA CAA-3'), at 200 pmol per six-well plate, by using Fugene6 (Roche). Whole-cell extracts and nuclear extracts were prepared 48 h after transfection. The silencing efficiency of *Irs1* and *Irs2* was determined by western blotting using antibodies specific for *Irs1* and *Irs2* (Cell Signaling).

Immunoblotting and Immunohistochemistry

We prepared cytosolic and nuclear protein extracts by SDS-PAGE and transferred them onto a nitrocellulose membrane (Schleicher & Schuell) by electroblotting. Foxa2 was

detected with antiserum against Foxa2 (diluted 1:1,000; a gift from R.I. Altaba). Foxo1 and Foxo1 phosphorylated on Ser 256 were detected with affinity-purified antibodies (1:1,000; Cell Signaling). Membranes were incubated with primary antibodies overnight at 4 °C.

Cryosections of livers (4 μ m) were fixed for 15 min at room temperature with 4% paraformaldehyde. For immunofluorescent detection of Foxa2 or haemagglutinin A (HA)-conjugated Foxa2, fixed sections were incubated overnight at 4 °C with antibodies against either Foxa2 (1:100) or HA (1:1,000; Covance). After washing, sections were treated with secondary antibody linked to Alexa Fluor 488 (Molecular Probes). For nuclear counterstaining, we used TOPRO-3 dye (Molecular Probes). Immunofluorescent staining was visualized by laser-scanning microscopy.

RT-PCR and Affymetrix gene array

Preparation of RNA, generation of probes for oligonucleotide expression arrays, hybridization conditions and RT-PCR are described in the Supplementary Information.

Laboratory measurements

Blood samples were taken from mice into non-heparinized capillary tubes. Insulin was quantified by a radioimmunoassay kit (Linco). Ketone bodies and free fatty acids were measured by a colorimetric assay system (Wako Chemicals). Glucose was measured by a standard glucose sensor (Glucometer Elite, Bayer). We determined cholesterol and triglycerides by a colorimetric assay system (Roche).

Liver perfusion

For the liver perfusion experiments, we used starved C57Bl/6J mice or *ob/ob* mice infected with either Ad-GFP or Ad-T156A (1×10^{11} particles per mouse) 7 d before the perfusion experiment. After anaesthesia with pentobarbitone sodium (60 mg per kg (body weight), intraperitoneum), the portal vein and the inferior vena cava were cannulated. The liver was perfused with oxygenated Krebs-Henseleit buffer with varying amounts of glucose and insulin at 37 °C in a single-pass mode with a total flow rate of 1.5 to 2 ml min⁻¹ (ref. 28). The outflow was collected and glucose concentrations were measured. Glucose output was calculated by subtracting the amount of glucose contained in the perfusion buffer from that measured in the outflow.

Mitochondrial β -oxidation

Mitochondria from perfused livers of mice were isolated by differential centrifugation as described²⁹. An aliquot of freshly isolated mitochondria was used to determine mitochondrial protein. We assessed the β -oxidation of [1-¹⁴C]palmitic acid by liver mitochondria as described³⁰. CO₂ trapped on the filter papers was counted for ¹⁴C activity by using a scintillation counter. The incubation mixture was centrifuged at 4,000g for 10 min and an aliquot of the supernatant was counted for ¹⁴C activity. This activity measures acid-soluble products of mitochondrial palmitate metabolism, which equals the formation of ketone bodies¹³.

Statistical analysis

Results are given as the mean $\pm$ s.d. Statistical analyses were done by a Student's *t*-test, and the null hypothesis was rejected at the 0.05 level. Linear regression was calculated by Origin (Microcal).

Received 29 July; accepted 20 September 2004; doi:10.1038/nature03047.

- Wolfrum, C., Besser, D., Luca, E. & Stoffel, M. Insulin regulates the activity of forkhead transcription factor Hnf-3 β /Foxa-2 by Akt-mediated phosphorylation and nuclear/cytosolic localization. *Proc. Natl Acad. Sci. USA* **100**, 11624–11629 (2003).
- Barthel, A. & Schmolli, D. Novel concepts in insulin regulation of hepatic gluconeogenesis. *Am. J. Physiol. Endocrinol. Metab.* **285**, E685–E692 (2003).
- Laffel, L. Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. *Diabetes Metab. Res. Rev.* **15**, 412–426 (1999).
- Casteels, K. & Mathieu, C. Diabetic ketoacidosis. *Rev. Endocr. Metab. Disord.* **4**, 159–166 (2003).
- Lewis, G. E., Carpentier, A., Adeli, K. & Giacca, A. Disordered fat storage and mobilization in the pathogenesis of insulin resistance and type 2 diabetes. *Endocr. Rev.* **23**, 201–229 (2002).
- McGarry, J. D. Banting lecture 2001: dysregulation of fatty acid metabolism in the etiology of type 2 diabetes. *Diabetes* **51**, 7–18 (2002).
- Nakae, J., Kitamura, T., Silver, D. L. & Accili, D. The forkhead transcription factor Foxo1 (Fkhr) confers insulin sensitivity onto glucose-6-phosphatase expression. *J. Clin. Invest.* **108**, 1359–1367 (2001).
- Franke, T. F., Kaplan, D. R., Cantley, L. C. & Tokier, A. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science* **275**, 665–668 (1997).
- Altomonte, J. et al. Inhibition of Foxo1 function is associated with improved fasting glycemia in diabetic mice. *Am. J. Physiol. Endocrinol. Metab.* **285**, E718–E728 (2003).
- Puigserver, P. et al. Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1 α interaction. *Nature* **423**, 550–555 (2003).
- Shimomura, I. et al. Insulin resistance and diabetes mellitus in transgenic mice expressing nuclear SREBP-1c in adipose tissue: model for congenital generalized lipodystrophy. *Genes Dev.* **12**, 3182–3194 (1998).
- Shimomura, I. et al. Decreased IRS-2 and increased SREBP-1c lead to mixed insulin resistance and sensitivity in livers of lipodystrophic and *ob/ob* mice. *Mol. Cell* **6**, 77–86 (2000).
- Freneaux, E. et al. Inhibition of the mitochondrial oxidation of fatty acids by tetracycline in mice and in man: possible role in microvesicular steatosis induced by this antibiotic. *Hepatology* **8**, 1056–1062 (1988).
- Weinstein, D. C. et al. The winged-helix transcription factor HNF-3 β is required for notochord development in the mouse embryo. *Cell* **78**, 575–588 (1994).
- Yoon, J. C. et al. Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).
- Herzig, S. et al. CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature* **413**, 179–183 (2001).

17. Matsuda, T., Noguchi, T., Yamada, K., Takenaka, M. & Tanaka, T. Regulation of the gene expression of glucokinase and L-type pyruvate kinase in primary cultures of rat hepatocytes by hormones and carbohydrates. *J. Biochem. (Tokyo)* **108**, 778–784 (1990).
18. Vaulont, S. & Kahn, A. Transcriptional control of metabolic regulation genes by carbohydrates. *FASEB J.* **8**, 28–35 (1994).
19. Chavin, K. D. *et al.* Obesity induces expression of uncoupling protein-2 in hepatocytes and promotes liver ATP depletion. *J. Biol. Chem.* **274**, 5692–5700 (1999).
20. An, J. *et al.* Hepatic expression of malonyl-CoA decarboxylase reverses muscle, liver and whole-animal insulin resistance. *Nature Med.* **10**, 268–274 (2004).
21. Voshol, P. J. *et al.* Increased hepatic insulin sensitivity together with decreased hepatic triglyceride stores in hormone-sensitive lipase-deficient mice. *Endocrinology* **144**, 3456–3462 (2003).
22. Stoffel, M. & Duncan, S. A. The maturity-onset diabetes of the young (MODY1) transcription factor HNF4 α regulates expression of genes required for glucose transport and metabolism. *Proc. Natl Acad. Sci. USA* **94**, 13209–13214 (1997).
23. Louet, J. F., Hayhurst, G., Gonzalez, F. J., Girard, J. & Decaux, J. F. The coactivator PGC-1 is involved in the regulation of the liver carnitine palmitoyltransferase I gene expression by cAMP in combination with HNF4 α and cAMP-response element-binding protein (CREB). *J. Biol. Chem.* **277**, 37991–38000 (2002).
24. Kim, H. I. & Ahn, Y. H. Role of peroxisome proliferator-activated receptor- γ in the glucose-sensing apparatus of liver and β -cells. *Diabetes* **53** (Suppl. 1), S60–S65 (2004).
25. Stumpfle, K. J., Koptides, M., Karinch, A. M. & Floros, J. Preparation of transcriptionally active nuclear extracts from mammalian tissues. *BioTechniques* **21**, 48–50, 52 (1996).
26. Unterman, T. G. *et al.* Hepatocyte nuclear factor-3 (HNF-3) binds to the insulin response sequence in the IGF binding protein-1 (IGFBP-1) promoter and enhances promoter function. *Biochem. Biophys. Res. Commun.* **203**, 1835–1841 (1994).
27. Shih, D. Q. *et al.* Hepatocyte nuclear factor-1 α is an essential regulator of bile acid and plasma cholesterol metabolism. *Nature Genet.* **27**, 375–382 (2001).
28. Tobey, T. A., Mondon, C. E., Zavaroni, I. & Reaven, G. M. Mechanism of insulin resistance in fructose-fed rats. *Metabolism* **31**, 608–612 (1982).
29. Hoppel, C., DiMarco, J. P. & Tandler, B. Riboflavin and rat hepatic cell structure and function. Mitochondrial oxidative metabolism in deficiency states. *J. Biol. Chem.* **254**, 4164–4170 (1979).
30. Lang, C. *et al.* Impaired hepatic fatty acid oxidation in rats with short-term cholestasis: characterization and mechanism. *J. Lipid Res.* **42**, 22–30 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Kruetzfeld for advice and comments. These studies were supported by grants from the NIH (M.S.), by an unrestricted grant from Bristol Myers Squibb (M.S.) and by the Howard Hughes Medical Institute (J.M.F.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (stoffel@rockefeller.edu).

The role of autophagy during the early neonatal starvation period

Akiko Kuma^{1,2,5,7}, Masahiko Hatano^{2,4}, Makoto Matsui^{5,6,7}, Akitsugu Yamamoto⁸, Haruaki Nakaya³, Tamotsu Yoshimori⁹, Yoshinori Ohsumi^{5,6}, Takeshi Tokuhisa² & Noboru Mizushima^{1,5,7}

¹Time's Arrow and Biosignaling, PRESTO, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan

²Department of Developmental Genetics (H2), ³Department of Pharmacology (F2), Chiba University Graduate School of Medicine, and ⁴Biomedical Research Center, Chiba University, Chiba 260-8670, Japan

⁵Department of Cell Biology, National Institute for Basic Biology, and

⁶Department of Molecular Biomechanics, School of Life Science, the Graduate University for Advanced Studies, Okazaki 444-8585, Japan

⁷Department of Bioregulation and Metabolism, Tokyo Metropolitan Institute of Medical Science, Tokyo 113-8613, Japan

⁸Department of Bio-Science, Nagahama Institute of Bio-Science and Technology, Nagahama 526-0829, Japan

⁹Department of Cell Genetics, National Institute of Genetics, Mishima 411-8540, Japan

At birth the trans-placental nutrient supply is suddenly interrupted, and neonates face severe starvation until supply can be restored through milk nutrients¹. Here, we show that neonates adapt to this adverse circumstance by inducing autophagy. Autophagy is the primary means for the degradation of cyto-

plasmic constituents within lysosomes^{2–4}. The level of autophagy in mice remains low during embryogenesis; however, autophagy is immediately upregulated in various tissues after birth and is maintained at high levels for 3–12 h before returning to basal levels within 1–2 days. Mice deficient for *Atg5*, which is essential for autophagosome formation, appear almost normal at birth but die within 1 day of delivery. The survival time of starved *Atg5*-deficient neonates (~12 h) is much shorter than that of wild-type mice (~21 h) but can be prolonged by forced milk feeding. *Atg5*-deficient neonates exhibit reduced amino acid concentrations in plasma and tissues, and display signs of energy depletion. These results suggest that the production of amino acids by autophagic degradation of 'self' proteins, which allows for the maintenance of energy homeostasis, is important for survival during neonatal starvation.

Autophagy is an intracellular, bulk degradation process in which a portion of cytoplasm is sequestered in an autophagosome and subsequently degraded upon fusion with a lysosome^{2–4}. Genetic studies on yeast have identified at least 16 *ATG* genes that are required for autophagosome formation⁵. Because autophagy-defective yeast mutants are not able to survive during nitrogen starvation⁶, autophagy is thought to be important for the cellular response to starvation, as well as the normal turnover of cytoplasmic constituents. Most of the *ATG* genes are conserved in higher eukaryotes. Mutations of the *ATG* genes in various species reveal a variety of phenotypes, such as: defective sporulation in *Saccharomyces cerevisiae*⁶, defective fruiting body formation in *Dictyostelium discoideum*⁷, premature death from the third larval to pupal stages in *Drosophila melanogaster*^{8,9}, and abnormal dauer formation in *Caenorhabditis elegans*¹⁰. In contrast, only minimal deficiencies (accelerated senescence) have been observed in plant *atg* mutants^{11,12}. Although many studies have suggested possible roles for autophagy in mammalian development, cell death and pathogenesis^{2,3}, genetic studies have been limited. *Atg6/Vps30*, which functions in at least two pathways in yeast (that is, autophagy and vacuolar protein sorting), has a mammalian orthologue called beclin 1 (*Becn1*). The *Becn1*^{-/-} mutation is lethal at embryonic day 7.5, and heterozygous mice (*Becn1*^{+/-}) exhibit increased tumorigenesis^{13,14}.

To study the physiological role of mammalian autophagy *in vivo*, we have generated a transgenic mouse model in which autophagosomes are labelled with GFP-LC3 in almost all tissues^{15,16}. LC3 is one of the mammalian proteins homologous to yeast *Atg8* (refs 17, 18). Using this mouse model, we have observed that autophagy is induced in many tissues in response to food withdrawal in young to adult mice¹⁵. We then extended this study to the embryonic and perinatal stages and observed that autophagy remained at a low level throughout the embryonic period. However, the formation of the GFP-LC3-labelled structures (GFP-LC3 'dots') that represent autophagosomes was extensively induced in various tissues after a natural birth. Particularly, the heart muscle, diaphragm, alveolar cells (Fig. 1a, b) and skin (not shown) displayed massive autophagy. Such an induction pattern is different from that of starved adult mice¹⁵. This might be because the energy requirements of the heart and diaphragm suddenly increase at birth, and the external environments of lung and skin are drastically changed; that is, from the amniotic fluid to the air. The appearance of autophagic vacuoles was confirmed by electron microscopy (Fig. 1c). Morphometric analysis of electron micrograph images revealed that autophagic vacuoles occupied 0.12% and 1.00% of the total cytoplasmic area in hearts isolated from neonates 0 h and 6 h after birth, respectively.

Furthermore, the induction of neonatal autophagy is immediate: formation of GFP-LC3 dots was upregulated within 30 min after birth (Fig. 1a, b). The autophagic activity reached its maximal level 3–6 h after birth, although the neonatal mice began suckling before that time. The number of GFP-LC3 dots gradually decreased to basal levels by day one or two. We then confirmed autophagy

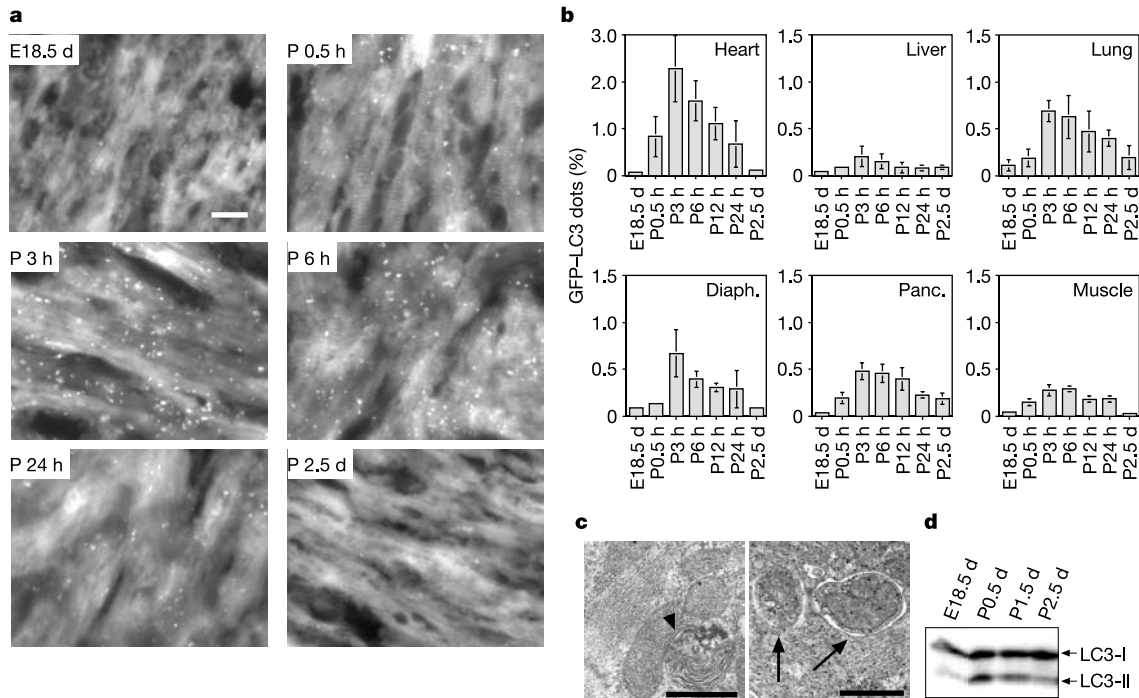


Figure 1 Autophagy is upregulated during the early postnatal period in wild-type mice. **a**, Autophagosome formation in the heart indicated by GFP–LC3. Hearts were isolated from GFP–LC3 transgenic mice at multiple stages, including 18.5 day embryos and 0.5 h, 3 h, 6 h, 24 h and 2.5 day neonates, immediately fixed, cryosectioned, and analysed by fluorescence microscopy. Scale bar, 10 μ m. **b**, Quantification of GFP–LC3 dots in neonatal tissues. Cryosections were prepared from tissues isolated at the indicated times. The ratio of the total area of GFP–LC3 dots to the total cellular area is shown as a

percentage. Values represent mean $\pm$ s.d. of three mice. Diaph., diaphragm; Panc., pancreas; Muscle; gastrocnemius muscle. **c**, Electron microscopic analysis of the heart from wild-type neonates. Typical autophagosomes (arrows) and autolysosomes (arrowhead) were observed in the heart at 10 h after birth. Scale bar, 1 μ m. **d**, Conversion of LC3-I to LC3-II in neonatal hearts. Heart homogenates from five embryos/neonates were pooled and analysed by immunoblotting using an anti-LC3 antibody. The bands representing LC3-I and LC3-II are marked with arrows.

induction by detection of the modification of endogenous LC3 using standard immunoblotting procedures^{16,17}. The amount of membrane-bound LC3-II (the phosphatidylethanolamine-conjugated form¹⁹) transiently increased on the day of birth and returned to basal levels by day two (Fig. 1d). These results suggest that massive autophagy is transiently induced in normal neonatal mice under physiological conditions, probably in response to the nutrient limitations imposed by the sudden termination of the trans-placental nutrient supply.

We postulated that neonates adapt to this unique kind of starvation by inducing autophagy to promote self-nourishment. We tested this hypothesis by generating *Atg5*^{−/−} mice. *Atg5* is an acceptor molecule for the ubiquitin-like molecule Atg12 (refs 20, 21). Our previous studies have demonstrated that the presence of *Atg5* and its proper conjugation with Atg12 are specifically required for the elongation of the autophagic isolation membrane²². *Atg5*^{+/-} embryonic stem (ES) cells were used to generate *Atg5*^{+/-} and *Atg5*^{−/−} mice (Fig. 2a). Although many studies have suggested possible roles for autophagy in development and cell death^{2,3}, *Atg5*^{−/−} mice were born at the expected mendelian frequency (+/+ : +/- : -/- = 150:371:147), and they appeared almost normal at birth (Fig. 2b). The body weight of *Atg5*^{−/−} mice (1.16 $\pm$ 0.137 g ($\pm$ s.d.), n = 13) was slightly lower than that of wild-type and heterozygous mice (1.29 $\pm$ 0.142 g, n = 50; P < 0.01). These data suggest that *Atg5*^{−/−} mice survive fetal development almost normally, which is in agreement with our observation that autophagic activity is generally low during the embryonic periods.

Neither the Atg12–Atg5 conjugate nor unconjugated Atg5 was detected in *Atg5*^{−/−} tissues and mouse embryonic fibroblasts (MEFs) (Fig. 2c). The amount of LC3-II was greatly reduced or

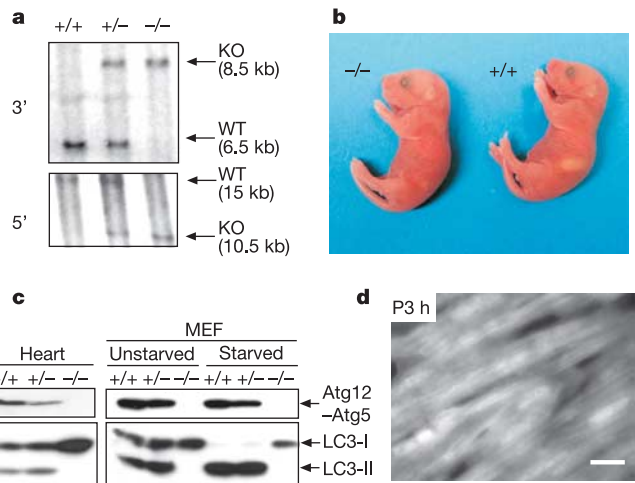


Figure 2 Generation of *Atg5*^{−/−} mice. **a**, Detection of wild-type (WT) and targeted alleles. Genomic DNA was prepared from mouse tails and analysed by Southern blot analysis. Restriction enzymes used were *HpaI* for the 5' probe and *SpeI* for the 3' probe (see Supplementary Fig.S2 for the restriction map). KO, knockout (targeted) allele. **b**, Photograph of a representative *Atg5*^{−/−} neonate compared to a wild-type littermate. **c**, Immunoblot analysis of Atg5 and LC3. Heart homogenates were prepared from neonates 10 h after birth. MEFs were cultured in Dulbecco's modified Eagle's medium supplemented with 10% FBS (unstarved) or Hanks' solution for 2 h (starved), and total cell lysates were prepared. Immunoblot analysis was performed using anti-Atg5 and anti-LC3 antibodies. The position of the Atg12–Atg5 conjugate on the blot is indicated. **d**, Absence of GFP–LC3 dot formation in *Atg5*^{−/−} mouse heart. Various tissues from *Atg5*^{−/−} mice expressing GFP–LC3 were analysed for dot formation by fluorescence microscopy. A representative result from a neonatal heart at 3 h after birth is shown.

absent in *Atg5*^{-/-} tissues and MEFs, as was previously observed in *Atg5*^{-/-} ES cells²². In contrast, the amount of LC3-I (the cytosolic form) was increased in *Atg5*^{-/-} tissues. Electron microscopic analysis confirmed that autolysosomes (degrading autophagic vacuoles) were not present in tissues from homozygous mutants (data not shown). When crossed with GFP-LC3 mice, the punctate GFP-LC3 dots that were observed in *Atg5*^{+/+} littermates were not generated in *Atg5*^{-/-} neonates (Fig. 2d). These data demonstrate that *Atg5*^{-/-} mice are indeed autophagy defective.

Despite the minimal abnormalities present at birth, most of the *Atg5*^{-/-} neonates died within 1 day of delivery. Of 52 *Atg5*^{-/-} mice, only one survivor was detected. This survivor died on the ninth day after birth. Heterozygous mice did not display any abnormal phenotypes during the extent of our monitoring period (that is, up to 16 months of age). Most of the *Atg5*^{-/-} neonates were found with no milk in their stomachs, suggesting that they may have had a suckling defect. However, histological examination of newborn *Atg5*^{-/-} mice revealed no obvious abnormalities, including the brain (Supplementary Fig. S1). To standardize the nutrient conditions, we compared the survival time of neonates under non-suckling conditions after caesarean delivery. *Atg5*^{+/+} and *Atg5*^{+/-} mice died at 20.6 ± 3.2 h (±s.d.) after birth, whereas *Atg5*^{-/-} mice died at 12.4 ± 1.3 h (*P* < 0.01) (Fig. 3a). Therefore, early death of the mutant homozygous mice was not simply due to suckling failure, indicating that autophagy has another crucial role for survival during this period. The survival time of *Atg5*^{-/-} neonates could be prolonged to more than 25 h by artificial milk feeding (Fig. 3b), suggesting that a major problem of *Atg5*^{-/-} neonates was a lack of nutrients. The fed *Atg5*^{-/-} mice finally died, probably because hand-feeding them milk was insufficient to combat their low nutrient status.

Because the major role of autophagy is the degradation of proteins into amino acids, we measured the plasma amino acid concentration under fasting conditions. Soon after the caesarean delivery, the concentration of amino acid in the plasma of *Atg5*^{-/-} neonates was not different from that of wild-type littermates (see Supplementary Table). However, at 10 h after the caesarean delivery, the total amino acid concentration of *Atg5*^{-/-} mice was significantly lower than that of wild-type mice (Fig. 4a). In particular, there were large differences in the plasma concentration of essential amino acids and branched-chain amino acids (BCAA). A similar pattern was observed for amino acid concentrations in tissues. In the organs that we examined (heart, liver and brain), the total amino acid concentration did not differ significantly among littermates at birth. However, after 10 h, significant differences were observed in those organs, particularly in the concentrations of essential amino acids and BCAAs (Fig. 4a and Supplementary Table). Therefore, early neonatal *Atg5*^{-/-} mice suffer from systemic amino acid insufficiency. Levels of other nutrients, as far as we examined, were not affected: the blood glucose level was less than 10 mg dl⁻¹ in both *Atg5*^{+/+} and *Atg5*^{-/-} mice, and the serum

free fatty acid concentration of *Atg5*^{-/-} mice (160 μequiv. l⁻¹) was as low as that of *Atg5*^{+/+} mice (148 μequiv. l⁻¹) 10 h after the caesarean delivery.

Our data demonstrating that neonatal mice exhibit severe hypoglycaemia and hypolipidaemia are in agreement with previous reports^{1,23}. Therefore, amino acids from degraded proteins might constitute the major determinant for energy metabolism until the nutrient supply from milk becomes steady. We then assessed the energy status of tissue by measuring the activity of AMP-activated protein kinase (AMPK), which functions as an energy sensor^{24,25}. AMPK is activated by the phosphorylation of the α-subunit of AMPK after an increase in the intracellular AMP:ATP ratio. A 10-h fasting treatment activated AMPK in the heart of neonatal *Atg5*^{-/-} mice but not of wild-type mice (Fig. 4b). Forced milk feeding of *Atg5*^{-/-} mice could suppress the AMPK activation (Fig. 4b, lane 7), confirming that the AMPK activation in *Atg5*^{-/-} mice is due to nutrient deficiencies. We next determined the amount of functional damage in the heart using electrocardiogram (ECG) recording. In

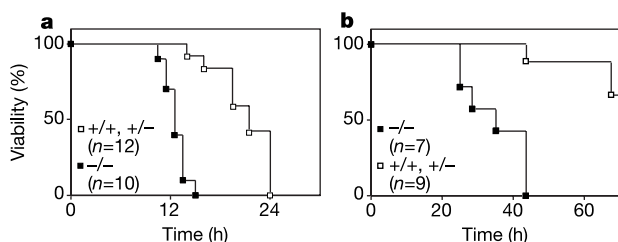


Figure 3 Early postnatal lethality of *Atg5*^{-/-} mice. Neonates were obtained by caesarean delivery. **a**, **b**, Successfully resuscitated pups were monitored in a humidified chamber without milk feeding (**a**) or with artificial milk feeding provided every 3–6 h through a tube inserted into the stomach (**b**).

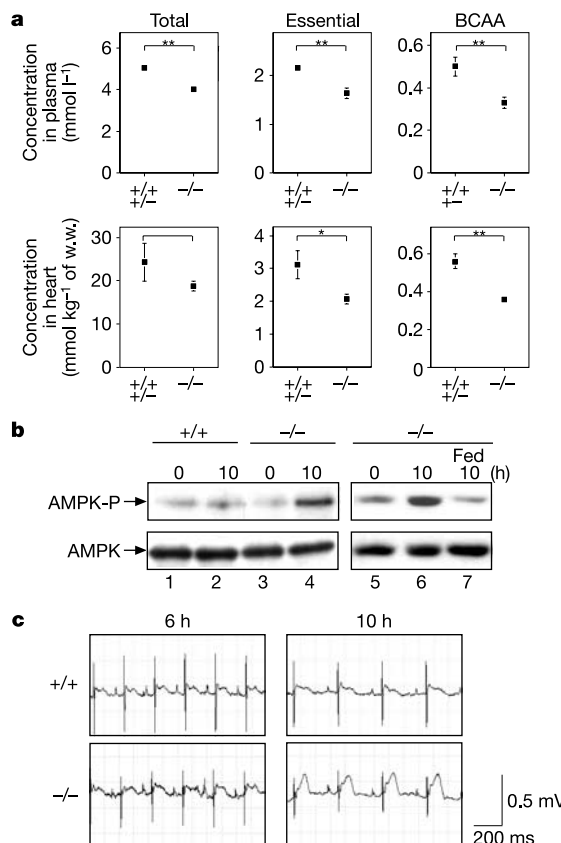


Figure 4 The energy depleted status of *Atg5*^{-/-} mice. **a**, Plasma and tissue amino acid concentrations. Amino acid concentrations were measured at 0 and 10 h after the caesarean delivery under fasting conditions. 'Total' indicates the sum of the Asp, Thr, Ser, Asn, Glu, Gln, Pro, Gly, Ala, Val, Cys, Met, Ile, Leu, Tyr, Phe, Lys, His and Arg concentrations; 'Essential' indicates the sum of Thr, Val, Met, Ile, Leu, Phe, Lys, His and Arg concentrations; 'BCAA' indicates the sum of the Val, Ile and Leu concentrations. Tissue amino acid concentrations are expressed as mmol kg⁻¹ of wet weight. Bars represent mean ± s.d. of three mice. Single and double asterisks indicate a significant difference between mutant and control (wild-type plus heterozygous) mice at *P* < 0.05 and *P* < 0.01, respectively. **b**, Activation of AMPK in *Atg5*^{-/-} mice. Hearts were isolated from unfed neonates at 0 h (lanes 1, 3, 5) and 10 h (lanes 2, 4, 6), and force-fed mice at 10 h (lane 7) after the caesarean delivery. Homogenates were prepared in the presence of phosphatase inhibitors and analysed by immunoblotting using anti-AMPK and anti-phospho-AMPK antibodies. **c**, ECGs from neonatal mice at the indicated time after caesarean delivery. ST elevation was observed in all *Atg5*^{-/-} mice (*n* = 4) at 8–9 h after delivery.

fasting *Atg5*^{-/-} mice 10 h after caesarean delivery, a severe elevation of the ST segment was observed; this elevation did not become apparent until 8 h after delivery (Fig. 4c). Fasting wild-type mice exhibited normal ECGs at 10 h, but a similar ST elevation appeared 3–5 h before death (data not shown). Although an ST elevation is usually suggestive of hypoxic damage of heart muscles, no abnormality was detected in the coronary arteries or the respiratory system of *Atg5*^{-/-} mice (data not shown). Therefore the ST elevation in those mice suggests a shortage of respiratory substrates, not oxygen. Taken together, these data suggest that neonates use the amino acids produced by autophagy for energy homeostasis.

We have demonstrated an important role for autophagy at early neonatal stages. Soon after birth, mammals encounter the first, and probably the most severe, period of starvation during their lifespan. To overcome this life-threatening problem unique to mammals, it has been known that carbohydrate and lipid reserves are used during this period¹. In addition, autophagy must be activated to maintain an adequate amino acid pool until the nutrient supply from milk reaches a steady state. Amino acids produced by autophagy can be directly used as an energy source or converted to glucose in the liver. Alternatively, amino acids can also be used for synthesis of proteins required for the proper starvation response. We could not prove that the nutrient supply derived from neonatal autophagy is essential for survival because of the presence of the potential suckling defect, which may partially account for the natural death of *Atg5*^{-/-} mice. Further analyses, particularly focusing on the central nervous system, will be required to disclose the importance of prenatal autophagy. Considering that the developmental defects of autophagy mutants reported in other species are related to nutrient starvation^{3,6–10}, the insufficiency of amino acids could explain many of these previously observed phenotypes that result from the loss of autophagy. □

Methods

Analysis of GFP–LC3 transgenic mice

Tissue samples for GFP–LC3 observation were prepared from late-stage embryos and fed neonates after natural delivery, and fixed with 4% paraformaldehyde as previously described¹⁵. Cryosections were imaged using a fluorescence microscope (Olympus IX81) equipped with a CCD camera (ORCA ER, Hamamatsu Photonics). To quantify the amount of GFP–LC3 dots, the dot signals were extracted using the Top Hat algorithm of Meta Morph Series version 6 (Molecular Device), and the total area of the dots was calculated. At least five independent visual fields from three mice were examined. All animal experiments were approved by the institutional committees of Tokyo Metropolitan Institute of Medical Science, Chiba University and National Institute for Basic Biology.

Electron microscopy

Mouse tissues were fixed with 2.5% glutaraldehyde in 0.1 M sodium phosphate buffer, pH 7.4, for 2 h, and electron microscopy was performed as previously described²⁶. Morphometric analysis was performed as previously described²².

Generation of *Atg5*^{-/-} mice

The targeted disruption of the *Atg5* gene of R1 ES cells has been previously reported²². *Atg5*^{+/-} ES cells were used to generate chimaeric mice using the aggregation method²⁷. Heterozygous mutant mice were outbred with C57BL/6 mice and interbred to obtain homozygous mutant mice. For histological examination, tissues were fixed in fresh 10% neutral buffered formalin and embedded in paraffin. Tissue sections were stained with haematoxylin and eosin, and examined by light microscopy. Embryonic fibroblasts were prepared from 13.5 day embryos. They were transformed with pEF321-T, an SV40 large T antigen expression vector (a gift from S. Sugano), and immortalized cell lines were established.

Southern blotting

Genomic DNA was digested with *HpaI* for the 5' probe and *SpeI* for the 3' probe. Next, it was separated by electrophoresis on an agarose gel, transferred to a nitrocellulose filter, and hybridized with probes of approximately 1 kilobase in length (shown in Supplementary Fig. S2), which had been labelled with digoxigenin by polymerase chain reaction as previously described²².

Western blotting

Mouse tissues were homogenized in nine volumes of ice-cold PBS supplemented with protease inhibitors. The homogenates were centrifuged at 500g for 10 min at 4 °C. Protein extracts (50 µg) were subjected to SDS–polyacrylamide gel electrophoresis and immunoblotting using anti-human *Atg5* (SO4) and anti-rat LC3 polyclonal antibodies

(SK2) that have been previously described^{17,22}. Phosphorylation of AMPK-α was determined by immunoblotting using an anti-phosphothreonine-172-specific antibody. Total AMPK was estimated using an anti-AMPK-α antibody (Cell Signaling Technology).

Caesarean delivery and artificial feeding

Pregnant female mice were injected with 2 mg of progesterone (luteum injection, Teikoku Hormone Mfg. Co.) on 17.5 days post coitus (d.p.c.) and 18.5 d.p.c. to delay birth. Newborn pups were obtained by caesarean delivery at 19.5 d.p.c. and placed in a humidified, thermostat-controlled chamber (30 °C). For artificial milk feeding, a fine tube was inserted into the stomach, and 30 µl of 0.13 mg ml⁻¹ infant formula for human neonates (Haihui, Wakodo Co.) was fed through the tube every 3–6 h. The final concentrations of carbohydrates, fat and protein in the formula were 7.1%, 3.6% and 1.6%, respectively.

Measurement of amino acid concentration

Plasma was de-proteinized with 3% trichloroacetate. Tissues were weighed while frozen and homogenized in glass micro-homogenizers with cold 5% sulphosalicylic acid. Free amino acids in the supernatants from both plasma and tissue samples were measured using an automated amino acid analyser (Hitachi L8500A).

ECG recordings

The surface ECG was recorded from newborn mice using a radio frequency transmitter (TA10ETA-F20, Data Science International), with subcutaneous leads placed in the conventional lead II position.

Received 10 August; accepted 17 September 2004; doi:10.1038/nature03029.

Published online 3 November 2004.

- Medina, J. M., Vicario, C., Juanes, M. & Fernandez, E. In *Perinatal Biochemistry* (eds Herrera, E. & Knopp, R.) 233–258 (CRC Press, Boca Raton, 1992).
- Cuervo, A. M. Autophagy: in sickness and in health. *Trends Cell Biol.* **14**, 70–77 (2004).
- Levine, B. & Klionsky, D. J. Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell* **6**, 463–477 (2004).
- Mizushima, N., Ohsumi, Y. & Yoshimori, T. Autophagosome formation in mammalian cells. *Cell Struct. Funct.* **27**, 421–429 (2002).
- Klionsky, D. J. *et al.* A unified nomenclature for yeast autophagy-related genes. *Dev. Cell* **5**, 539–545 (2003).
- Tsukada, M. & Ohsumi, Y. Isolation and characterization of autophagy-defective mutants of *Saccharomyces cerevisiae*. *FEBS Lett.* **333**, 169–174 (1993).
- Otto, G. P., Wu, M. Y., Kazgan, N., Anderson, O. R. & Kessin, R. H. Macroautophagy is required for multicellular development of the social amoeba *Dictyostelium discoideum*. *J. Biol. Chem.* **278**, 17636–17645 (2003).
- Juhász, G., Csikos, G., Sinka, R., Erdélyi, M. & Sass, M. The *Drosophila* homolog of Aut1 is essential for autophagy and development. *FEBS Lett.* **543**, 154–158 (2003).
- Scott, R. C., Schuldiner, O. & Neufeld, T. P. Role and regulation of starvation-induced autophagy in the *Drosophila* fat body. *Dev. Cell* **7**, 167–178 (2004).
- Melendez, A. *et al.* Autophagy genes are essential for dauer development and life-span extension in *C. elegans*. *Science* **301**, 1387–1391 (2003).
- Doelling, J. H., Walker, J. M., Friedman, E. M., Thompson, A. R. & Veirstra, R. D. The APG8/12-activating enzyme APG7 is required for proper nutrient recycling and senescence in *Arabidopsis thaliana*. *J. Biol. Chem.* **277**, 33105–33114 (2002).
- Hanaoka, H. *et al.* Leaf senescence and starvation-induced chlorosis are accelerated by the disruption of an *Arabidopsis* autophagy gene. *Plant Physiol.* **129**, 1181–1193 (2002).
- Yue, Z., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl Acad. Sci. USA* **100**, 15077–15082 (2003).
- Qu, X. *et al.* Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. *J. Clin. Invest.* **112**, 1809–1820 (2003).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Mizushima, N. Methods for monitoring autophagy. *Int. J. Biochem. Cell Biol.* **36**, 2491–2502 (2004).
- Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J.* **19**, 5720–5728 (2000).
- Ichimura, Y. *et al.* A ubiquitin-like system mediates protein lipidation. *Nature* **408**, 488–492 (2000).
- Kabeya, Y. *et al.* LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J. Cell Sci.* **117**, 2805–2812 (2004).
- Mizushima, N. *et al.* A protein conjugation system essential for autophagy. *Nature* **395**, 395–398 (1998).
- Mizushima, N., Sugita, H., Yoshimori, T. & Ohsumi, Y. A new protein conjugation system in human. The counterpart of the yeast Apg12p conjugation system essential for autophagy. *J. Biol. Chem.* **273**, 33889–33892 (1998).
- Mizushima, N. *et al.* Dissection of autophagosome formation using Apg5-deficient mouse embryonic stem cells. *J. Cell Biol.* **152**, 657–667 (2001).
- Brun, S. *et al.* Activators of peroxisome proliferator-activated receptor-α induce the expression of the uncoupling protein-3 gene in skeletal muscle: a potential mechanism for the lipid intake-dependent activation of uncoupling protein-3 gene expression at birth. *Diabetes* **48**, 1217–1222 (1999).
- Hardie, D. G. Minireview: the AMP-activated protein kinase cascade: the key sensor of cellular energy status. *Endocrinology* **144**, 5179–5183 (2003).
- Carling, D. The AMP-activated protein kinase cascade—a unifying system for energy control. *Trends Biochem. Sci.* **29**, 18–24 (2004).
- Yamamoto, A. *et al.* Stacks of flattened smooth endoplasmic reticulum highly enriched in inositol 1,4,5-trisphosphate (InsP3) receptor in mouse cerebellar Purkinje cells. *Cell Struct. Funct.* **16**, 419–432 (1991).

27. Wood, S. A., Allen, N. D., Rossant, J., Auerbach, A. & Nagy, A. Non-injection methods for the production of embryonic stem cell-embryo chimaeras. *Nature* **365**, 87–89 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Miwa and H. Satake for technical assistance. We also thank S. Sugano for donation of the pEF321-T plasmid; K. Ono and K. Tanaka for histological examination of the brain; M. Tamagawa for instruction in electrocardiogram recording; and S. Nishio, N. Tsunekawa and M. Terai for discussions. Amino acid measurements were carried out with the aid of the Center for Analytical Instruments at the National Institute for Basic Biology. This work was supported in part by Grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.M. (nmizu@rinshoken.or.jp).

An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division

Ralf Kittler¹, Gabriele Putz¹, Laurence Pelletier¹, Ina Poser¹, Anne-Kristin Heninger¹, David Drechsel¹, Steffi Fischer¹, Irena Konstantinova¹, Bianca Habermann², Hannes Grabner¹, Marie-Laure Yaspo³, Heinz Himmelbauer³, Bernd Korn⁴, Karla Neugebauer¹, Maria Teresa Pisabarro^{1*} & Frank Buchholz¹

¹Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, D-01307 Dresden, Germany

²Scionics Computer Innovation, GmbH, Pfotenhauerstrasse 110, D-01307 Dresden, Germany

³Max Planck Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany

⁴RZPD-Ressourcenzentrum für Genomforschung, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany

* Present address: TU Dresden, Biotechnologisches Zentrum, Tatzberg 47-51, D-01307 Dresden, Germany

RNA interference (RNAi) is an evolutionarily conserved defence mechanism whereby genes are specifically silenced through degradation of messenger RNAs; this process is mediated by homologous double-stranded (ds)RNA molecules^{1–4}. In invertebrates, long dsRNAs have been used for genome-wide screens and have provided insights into gene functions^{5–8}. Because long dsRNA triggers a nonspecific interferon response in many vertebrates, short interfering (si)RNA or short hairpin (sh)RNAs must be used for these organisms to ensure specific gene silencing^{9–11}. Here we report the generation of a genome-scale library of endoribonuclease-prepared short interfering (esi)RNAs¹² from a sequence-verified complementary DNA collection representing 15,497 human genes. We used 5,305 esiRNAs from this library to screen for genes required for cell division in HeLa cells. Using a primary high-throughput cell viability screen followed by a secondary high content videomicroscopy assay, we identified 37 genes required for cell division. These include several splicing factors for which knockdown generates mitotic spindle defects. In addition, a putative nuclear-export terminator was found to speed up cell proliferation and mitotic progression after knockdown. Thus, our study uncovers new aspects of cell division and establishes esiRNA as a versatile approach for genomic RNAi screens in mammalian cells.

We generated a large-scale RNAi library using an alternative approach to previously reported RNAi expression libraries^{13–15}. Our approach was based on the processing of long dsRNA by *Escherichia coli* RNaseIII *in vitro* (Fig. 1a). EsiRNAs generate a variety of siRNAs, which are able to efficiently and specifically silence target mRNA; this abolishes the need to identify effective silencers for each mRNA^{16–18}. Moreover, this technology, compared with transfection- or viral expression-based approaches, allows far greater control over interfering dsRNAs at the cellular level; low rates of plasmid transfection or variations in virus titres can be problematic for RNAi studies. Because of its simplicity, efficiency and cost-effectiveness (approximately US\$4 per gene), esiRNA has advantages for large-scale loss-of-function studies in mammalian cells.

Assuming that genes essential for cell division are likely to present a cell viability phenotype, we used the simple and fast WST-1 assay as an initial screen to identify 275 genes; these genes were then studied in detail by videomicroscopy to allow detailed spatial and temporal visualization of mitosis and cytokinesis (Fig. 1b). We observed severe cell division phenotypes for 37 genes (Table 1 and Supplementary Table 1). Each of these genes was checked for functional annotation in the HARVESTER¹⁹ unification database and in the scientific literature. No functional annotation was found for seven out of the 37 genes, so we can assign a function in cell division to seven previously uncharacterized genes. For example, knockdown of the predicted mRNAs DKFZP564M082 and FLJ30851 resulted in mitotic arrest and severe spindle defects in the cells (Figs 2d and 3g, h). Functional annotations were found for the remaining 30 genes, but seven of these are based on electronic annotation and have not been experimentally verified. Interestingly, 23 genes had previously only been associated with functions other than cell division.

We grouped the 37 observations of cell division phenotypes (<http://www.mpi-cbg.de/esiRNA>) into three categories: mitotic arrest (Fig. 2b, d and Supplementary Fig. 1), aberrant cytokinesis (Fig. 2g, h) and cell death upon entry into mitosis (Fig. 2i). Nineteen out of 37 esiRNAs that caused mitotic arrest led to cell death after prolonged latency in mitosis. These genes probably stop the cells from progressing further through mitosis and ultimately induce cell death. Six of the 37 mRNA knockdowns caused the cells to die quickly upon entry into mitosis. Therefore, these genes might be directly linked to cell death at the point of entry into mitosis.

Twelve esiRNAs led to an aberrant cytokinesis phenotype, of which ten also displayed a mitotic arrest phenotype. These phenotypes could further be divided into two classes. The knockdown of *ITPR1*, *MFAP1*, *AD024*, *GALNT5*, *CKLFSF4* and *KIAA0056* resulted in a cell cleavage defect leading to bi-nucleated cells after mitosis (Fig. 2g). A different cytokinesis phenotype was observed for *SNRPA1*, *SNRPB*, *DHX8*, *SNW1*, *FLJ10290*, and importin- β (*KPNB1*). With these knockdowns, some of the cells become round (as normally observed for cells entering mitosis) but exit mitosis some time later, without any sign of cell division. These cells then contain only one nucleus, which has a disorganized appearance; for example, substructures such as the nucleoli are not visible (Fig. 2h). For *DHX8* and *SNW1*, we also observed a cytokinesis defect that led to the formation of cell fragments devoid of chromatin (cytoplasm) (Fig. 3k, l).

We further examined the cell division phenotypes by analysing spindle morphology after mRNA knockdown. We observed severe spindle defects for 23 genes displaying different aberrations (Table 1). The RNAi phenotype of 20 genes showed spindles with two or more microtubule foci and a reduced number of microtubule connections between the foci and the chromosomes (Fig. 3d–h), indicating a defect in microtubule assembly and/or the centrosome cycle. For seven genes displaying mitotic arrest upon mRNA knockdown, we observed no obvious spindle defects, indicating that a different primary defect must be associated with the mitotic arrest.

The knockdown of *KIF11* (Fig. 3b) and the nuclear RNA helicase

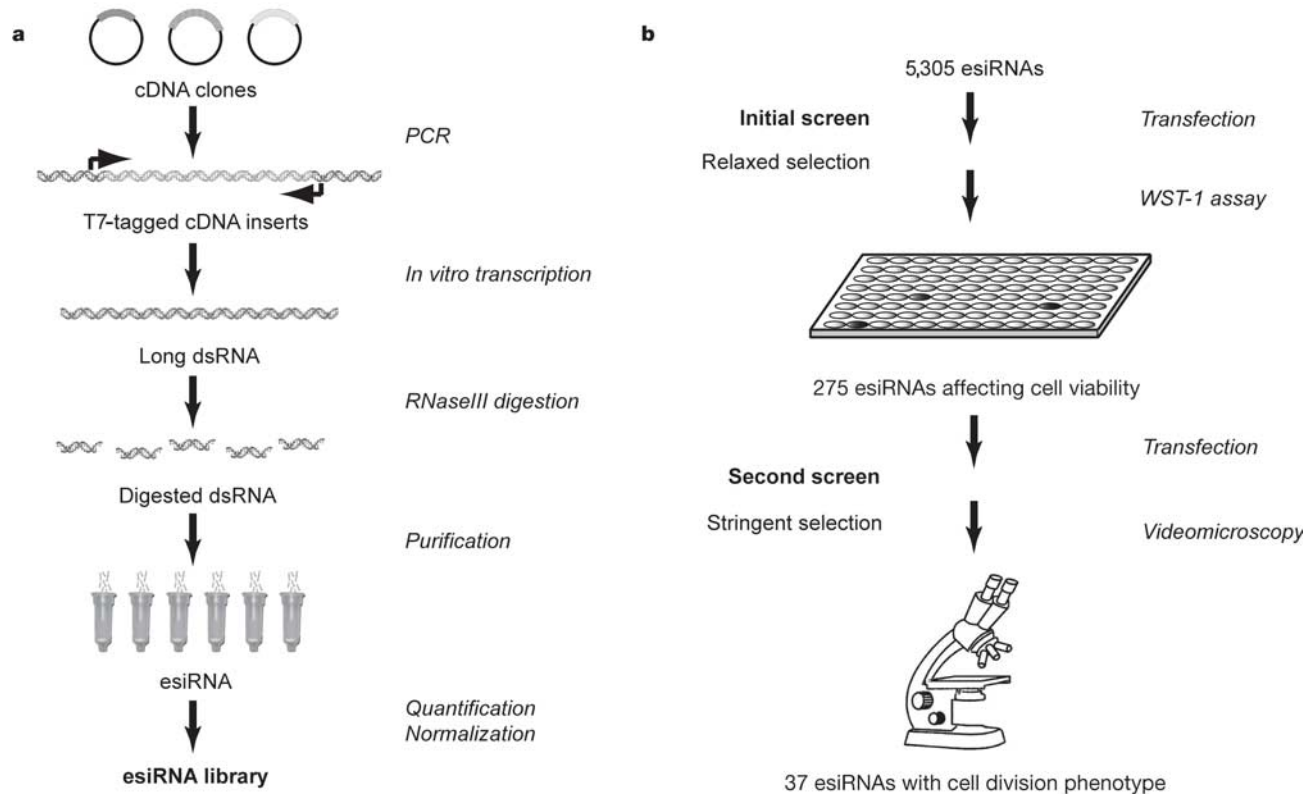


Figure 1 Library generation and screening strategy. **a**, Scheme for the generation of a large-scale esiRNA library. All steps were done in 96-well plates. **b**, Strategy for screening genes essential for cell division. Important steps are shown in *italic*.

Table 1 Thirty-seven genes essential for cell division in human cells

Gene	Functional annotation	Mitotic arrest	Cytokinesis defect	Death*	Spindle
<i>CDC16</i>	APC/C subunit	Yes	No	No	Abnormal
<i>CDC27</i>	APC/C subunit	Yes	No	No	Abnormal
<i>CENPE</i>	Motor protein	Yes	No	No	Normal
<i>KIF11</i>	Motor protein	Yes	No	No	Abnormal
<i>KIAA0056</i>	Condensin subunit	Yes	Yes	No	Abnormal
<i>AD024</i>	Kinetochore protein	Yes	Yes	No	Abnormal
<i>RRM2</i>	Ribonucleotide reductase	No	No	Yes	Abnormal
<i>GALNT5</i>	Glycosyl transferase	Yes	Yes	No	Normal
<i>CKLFSF4</i>	Cytokine	No	Yes	No	Normal
<i>HRI</i>	Protein kinase	Yes	No	No	Abnormal
<i>ITPR1</i>	Receptor protein	No	Yes	No	Normal
<i>PSCD3</i>	Receptor protein	Yes	No	No	Normal
<i>CASP8AP2</i>	Apoptosis factor	No	No	Yes	Normal
<i>DDX5</i>	Splicing factor	Yes	No	No	Abnormal
<i>DHX8</i>	Splicing factor	Yes	Yes	No	Abnormal
<i>LSM6</i>	Splicing factor	Yes	No	No	Abnormal
<i>SART1</i>	Splicing factor	Yes	No	No	Abnormal
<i>SNRPA1</i>	Splicing factor	Yes	Yes	No	Abnormal
<i>SNRPB</i>	Splicing factor	Yes	Yes	No	Abnormal
<i>SNW1</i>	Splicing factor	Yes	Yes	No	Abnormal
<i>MFAP1</i>	Microfibril component	Yes	Yes	No	Abnormal
<i>ZDHHC5</i>	Transcription factor	Yes	No	No	Normal
<i>DDX48</i>	Translation factor	Yes	No	No	Abnormal
<i>EIF3S3</i>	Translation factor	No	No	Yes	Unclear
<i>EIF3S10</i>	Translation factor	No	No	Yes	Unclear
<i>CLDN16</i>	Structural protein	Yes	No	No	Abnormal
<i>KPNB1</i> (importin- β)	Importin	Yes	Yes	No	Abnormal
<i>VCP</i> (p97)	Vesicular transport protein	Yes	No	No	Abnormal
<i>AP1S1</i>	Vesicular transport protein	Yes	No	No	Normal
<i>ATP6V1D</i>	Vacuolar ATPase	Yes	No	No	Normal
<i>DKFZP564M082</i>	Unknown	Yes	No	No	Abnormal
<i>FLJ10290</i>	Unknown	Yes	Yes	No	Unclear
<i>FLJ20311</i>	Unknown	Yes	No	No	Abnormal
<i>FLJ30851</i>	Unknown	Yes	No	No	Abnormal
<i>MGC3248</i>	Unknown	Yes	No	No	Normal
<i>FLJ38663</i>	Unknown	No	No	Yes	Abnormal
<i>MGC2603</i>	Unknown	No	No	Yes	Normal

* Cell death upon entry into mitosis.

DDX48 (Fig. 3c) resulted in monopolar spindles. *KIF11* is a kinesin-like motor protein that is required for centrosome separation and therefore the establishment of a bipolar spindle^{20,21}. For RNAi of importin- β we observed many rounded-up cells with apparently uncondensed chromatin and a level of microtubule nucleation consistent with interphase (Fig. 3i). Thus, although the cell cortex displays the characteristics of a mitotic cell, the presence of uncondensed chromatin suggests that these cells are in interphase. Because importin- β is a key component of Ran-dependent nuclear protein import²², the RNAi phenotype suggests that importin- β is required for the import of proteins that trigger the progression of the nucleus from interphase into mitosis.

Surprisingly, many genes that displayed cell division phenotypes with spindle defects are known components of the splicing machinery, including *SNRPA1*, *SNRPB*, *SNW1*, *DHX8*, *DDX5*, *LSM6* and *SART1*. In principle, there are two possible explanations for this phenomenon. The simplest explanation would be that the cell division defect is an indirect consequence of defects in pre-mRNA splicing. In this scenario, the depletion of splicing factors leads to impaired pre-mRNA processing of genes essential for cell division processes, in turn resulting in depletion of the corresponding protein(s). This has been shown for the temperature-sensitive yeast splicing factors *PRP16*, *PRP19* and *Cef1*, which impair α -tubulin synthesis thereby preventing spindle assembly^{23,24}. According to this scenario, the knockdown of splicing factors precedes the depletion of gene(s) essential for cell division, and impaired mitosis is a resulting secondary effect.

Alternatively, some splicing factors might have a direct role in cell division. Several points indicate that this might be the case. First, an indirect effect of splicing factors on cell division would imply that

the turnover rate of these proteins and the subsequent depletion of mitosis-specific gene(s) has to be fast. However, it has recently been reported that most splicing factors have a low turnover rate²⁵. Because all spliceosome RNAi phenotypes became visible 36–48 h after transfection, it is difficult to imagine that the subsequent depletion of mitosis-specific gene(s) would be fast enough and sufficient to explain the cell division phenotypes. Second, in the yeast experiments mentioned above, a marked reduction of α -tubulin levels was noted upon inactivation of spliceosome subunits^{23,24}. In contrast, our tubulin stainings did not show a detectable change in α -tubulin levels (Fig. 3). Third, nuclear run-on experiments of several splicing factor knockdowns showed that general transcription, and therefore splicing, is not affected (data not shown). Fourth, *TPX2*, which is required for chromosome-induced microtubule assembly during spindle formation²⁶, has recently been shown to co-purify with spliceosome components²⁷. Interestingly, the spindle defect observed upon *TPX2* depletion by RNAi resembles the phenotype observed here with some splicing factors; this supports the idea of a direct link between splicing and cell division. In this context it is worth mentioning that knockdown of *MFAP1*, an extracellular matrix component of the elastin-associated microfibrils, results in a cell division phenotype similar to that observed with the splicing factors in our study; mass spectrometry also showed *MFAP1* to be a component of the spliceosome²⁷. Future studies will be required to determine whether the observed cell division defects result from direct or indirect interactions between splicing factors and the spindle assembly machinery.

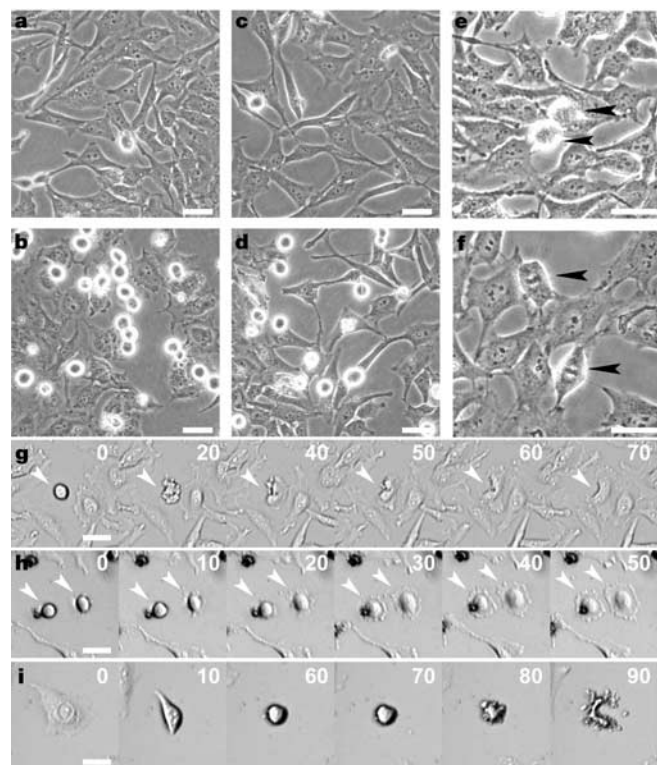


Figure 2 Cell division phenotypes visualized by videomicroscopy. **a–i**, HeLa cells were transfected with esiRNA targeting *SNW1* (**b**), *FLJ30851* (**d**), *KIAA1387* (**f**), firefly luciferase (negative control, **a**, **c**, **e**), *GALNT5* (**g**), *SNRPA1* (**h**) or *RRM2* (**i**). Microscopic images were taken after 48 h (**a**, **b**, **c**, **d**) or 72 h (**e** and **f**). The arrows indicate cells in metaphase (**e**), the *KIAA1387*-knockdown phenotype (**f**) and cells with cytokinesis defects (**g**, **h**). The arrows in **g** indicate the generation of two nuclei in a single cell. Numbers indicate minutes after the start of the time-lapse sequence. Scale bars are 40 μ m.

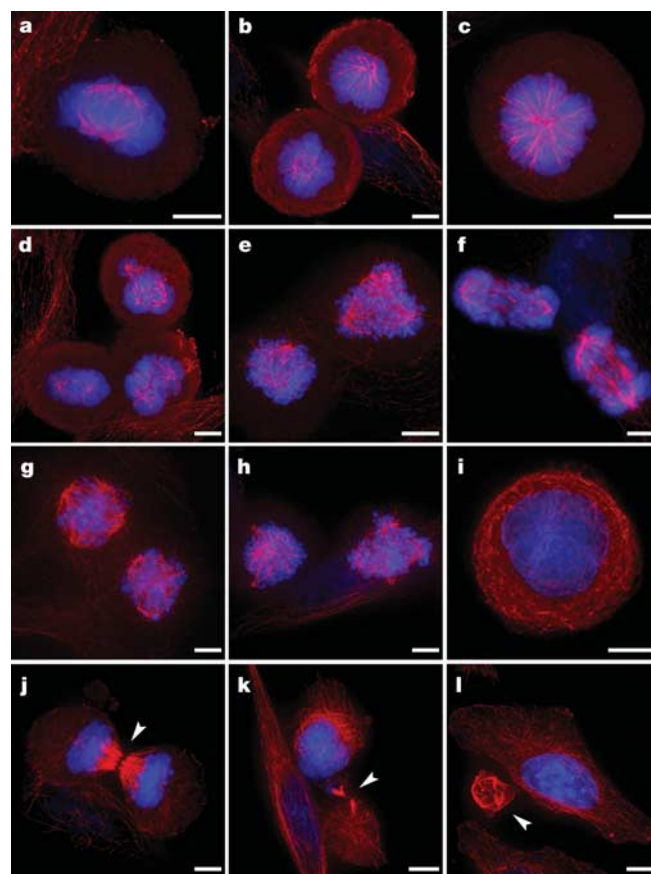


Figure 3 Spindle and cytokinesis defects observed for RNAi phenotypes. **a–l**, HeLa cells were transfected with esiRNA targeting firefly luciferase (negative control, **a**, **j**), *KIF11* (**b**), *DDX48* (**c**), *SNW1* (**d**, **l**), *DHX8* (**e**, **k**), *MFAP1* (**f**), *DKFZP564M082* (**g**), *FLJ30851* (**h**), importin- β (**i**) and imaged by three-dimensional deconvolution microscopy 48 h after transfection. Cells were stained for tubulin (red) and DNA (blue). Scale bars are 5 μ m. The arrows in **j** and **k** indicate the position of the cleavage furrow. The arrow in **l** points to the collapsed cytoplasm.

Another unexpected finding of this study was the identification of a gene for which knockdown increases cell growth rate. So far, loss-of-function leading to an increase in cell growth has only been identified for a very small number of genes. One such gene, *P27KIP1*, opposes mitotic stimuli and therefore has a negative regulatory role on cell proliferation²⁸. Consistent with this, RNAi of *P27KIP1* resulted in a significantly increased growth rate in our screen (data not shown). For another gene in this class (*KIAA1387*), we observed a remarkable cell division phenotype of continuous cellular attachment during mitosis (Fig. 2f). The partial detachment or 'rounding-up' of the cell during mitosis is a general feature of many epithelial cells, although the reason for this morphological change is not clear. In addition, cells with RNAi for *KIAA1387* also progressed significantly faster through mitosis with a transit time of 26 ± 7 min compared to 36 ± 8 min for control cells. Because the complete sequence of human *KIAA1387* was not known, we cloned and sequenced this gene from cDNA, and identified two splice variants in HeLa cells. Using green fluorescent protein (GFP) fusion constructs, we determined that *KIAA1387* localizes to the nucleus (see Supplementary Fig. 2a–d). Bioinformatic analyses suggest that *KIAA1387* contains a Ran-binding domain, a DUF625 domain (of unknown function) and ARM (armadillo) repeat domains (see Supplementary Fig. 2e). On the basis of these analyses, we postulate that *KIAA1387* is a nuclear, Ran-binding protein implicated in nuclear–cytoplasmic transport processes (see Supplementary Information). Studies are currently underway to investigate how this role results in the observed phenotype.

Crucial parameters for RNAi screens are the efficiency and specificity with which each gene is silenced. In order to test the efficiency of our screen we performed three analyses. First, we chose two multi-protein complexes, the 20S (core) proteasome and the ribosome, and identified which of their subunit genes were present in the esiRNA library after the cell viability screen. As both of these

complexes are essential, knockdown of these proteins should have an effect in the cell viability assay. A similar strategy was recently used to validate an shRNA library, which identified five out of twelve core subunits essential for proteasome function¹⁵. In our study, nine out of nine proteasome core subunits (Fig. 4a) and 23 out of 26 ribosomal proteins (see Supplementary Table 2 and Supplementary Fig. 3) were identified using the cell viability assay. Importantly, the three remaining ribosomal genes either were not expressed in HeLa cells (the Y-chromosome-specific isologue *RPS4Y*) or are retroposons, which are dispensable for ribosomal structure and function (Fig. 4b). Therefore, we detected all of the genes from these protein complexes that are essential for growth and viability. Second, we compared published data on RNAi-mediated cell division phenotypes in tissue culture cells with our data set. We found 26 publications reporting 32 cell division defects upon mRNA knockdown. Of these, 11 genes were present in our esiRNA library, and nine of these were reported to result in reduced cell viability (see Supplementary Table 3). Six of these nine genes were picked up in our cell viability assay. Third, the percentage of essential genes detected by our screen in HeLa cells is almost identical to that reported recently for *Drosophila melanogaster* tissue culture cells⁸. Further analysis revealed that in our esiRNA library, 71.8% of the human orthologues to essential fruitfly genes were also found to be essential for cell viability in HeLa cells (see Supplementary Information). This probably reflects conservation of fundamental biological processes between fly and human cells, and indicates that the esiRNA screen in HeLa cells was similarly efficient to the screen in *Drosophila* cells.

In order to evaluate the specificity of esiRNA in our experiments, we first analysed the set of ribosomal genes in more detail. The three ribosomal genes that were not identified in the screen share 79–89% DNA sequence similarity to genes that were identified (*RPS4X/RPS4Y*, *RPL12/RPL12-like*, *RPL10/RPL10-like*), suggesting that esiRNA is highly target-specific. This was further confirmed by quantification of *RPS4X* knockdown after transfection of esiRNAs specific for *RPS4X* and *RPS4Y*, which are 82% identical (Fig. 4c and Supplementary Fig. 4). Second, we generated independent esiRNAs for 21 genes identified as essential for cell division, using a different region of each of the genes for esiRNA production. All of the new esiRNAs produced the same phenotype as the original esiRNAs (data not shown), suggesting a high degree of specificity for esiRNAs. This is not unexpected, because pooling of siRNAs (esiRNA represents a heterogeneous pool of siRNAs targeting the same mRNA) has been suggested to reduce off-target effects^{29,30}.

Our screen has uncovered additional human genes associated with cell division, thereby creating a starting point for further insight into this fundamental biological process. Our work also highlights the potential of using esiRNA in mammals to screen individual genes. The method is efficient, specific and easily adjustable to different cell-based assays, and is therefore well suited to provide insight into various aspects of mammalian biology. □

Methods

Generation of the esiRNA library

We generated a library from cDNA clones of the Human Ensembl Set, which were obtained from the German Genome Resource Center (RZPD). These clones have been sequence-verified in order to ensure correct annotation. One advantage of using the Ensembl Set is that these clones are derived from inserts with a length of 500–1,200 base pairs (bp), close to the 3' end of the transcript. This avoids highly conserved or repetitive sequences (the Ensembl Set was originally created for the production of spot arrays for gene expression analyses). This clone set is well suited to the synthesis of esiRNAs because the exclusion of highly conserved sequences reduces the risk of cross-silencing homologous genes. Detailed protocols for tagging cDNA clone inserts with T7 promoter sequences (using PCR), *in vitro* transcription, digestion of long dsRNA and purification of esiRNA can be found in the Supplementary Information. Details on esiRNA library generation can also be found at <http://www.mpi-cbg.de/esirna>.

Cell-based RNAi screen

HeLa cells were seeded 16 h before transfection into 96-well plates at a density of 1,200 cells

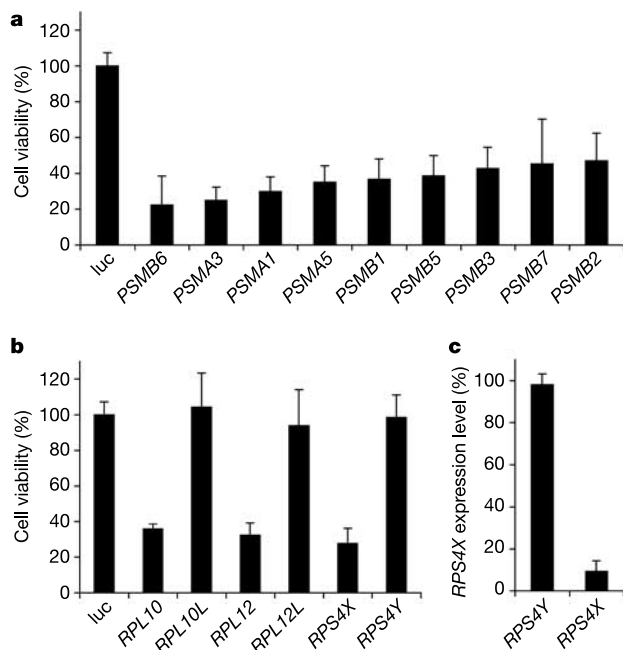


Figure 4 Efficiency and specificity of esiRNA. **a**, Effect on cell viability of esiRNAs targeting nine different subunits of the 20S proteasome. **b**, Effect on cell viability of three esiRNAs targeting essential ribosomal proteins (*RPL10*, *RPL12* and *RPS4X*) and their paralogs (*RPL10L*, *RPL12L* and *RPS4Y*). Cells in **a** and **b** were assayed 72 h after transfection. Cell viability was normalized against control esiRNA transfected cells (esiRNA directed against firefly luciferase, luc). **c**, *RPS4X* mRNA expression after transfection of esiRNAs targeting *RPS4X* and *RPS4Y* (see also Supplementary Fig. 4). All error bars represent s.d.

per well in 100 μ l medium (DMEM, 10% FBS, 2 mM L-glutamine, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin). Transfections were performed in triplicate using 25 ng esiRNA and oligofectamine (Invitrogen). Cell viability was measured 72 h after transfection using the WST-1 assay (Roche Diagnostics) with a DigiScan 400 microplate reader (ASYS) at 450 nm, using 620 nm as the reference wavelength. The background was subtracted from the absorption readings and the values were mean-centred for each 96-well plate. In the initial screen we typically selected all genes for which a normalized value greater than 1.645 s.d. or less than -1.645 s.d. was obtained at least twice. These cutoff values represent 10% of all outliers, assuming a normal distribution of values. For all positive hits we re-synthesized esiRNA starting from the original cDNA clones, which were re-sequenced and confirmed by BLAST search against the published sequence of the human genome.

We re-arrayed the esiRNAs for the proteasome core subunits, all 26 cytosolic ribosomal proteins and negative controls on a 96-well plate. Transfection and WST-1 assay were performed as described above.

Time-lapse microscopic assay

For the secondary screen we re-arrayed the esiRNAs that scored in the first screen with three negative controls (esiRNA targeting firefly luciferase) per 96-well plate. We examined all 275 genes using time-lapse bright-field microscopy, capturing one frame every 10 min over 48 h, starting 30 h after transfection. This was done with an Axiovert 200 microscope (Zeiss), using a 5 $\times$ objective with an integrated automatic stage and incubator using METAMORPH software. We quantified the frequency of cells in mitosis for all arrest phenotypes (Supplementary Fig. 1) by counting an average of 200 cells at three different time points (frame 1, frame 150 and frame 251). For KIAA1387 and the luciferase control, the average length of time in mitosis was calculated for 40 cells by counting the number of frames between entry into and exit from mitosis from frame 250 onward.

Immunofluorescence and microscopy

For visualizing the morphology of the mitotic spindle, HeLa cells that had been transfected and grown on coverslips were fixed and permeabilized in methanol at -20°C for 8 min. Cells were washed with PBS and incubated for 10 min in PBS containing 0.2% fish skin gelatine (PBS/FSG, Sigma) to prevent nonspecific binding of antibodies. Cells were incubated for 20 min with a mouse monoclonal antibody against tubulin (DM1, Sigma) diluted to 1 μ g ml⁻¹ in PBS/FSG, washed in PBS, incubated in 1 μ g ml⁻¹ Texas-red-conjugated secondary antibodies for 20 min at 37°C, and washed in PBS before mounting a coverslip in the presence of DAPI (1 μ g ml⁻¹) to visualize chromatin. Three-dimensional data sets were acquired using a DeltaVision imaging system (Applied Precision) equipped with an Olympus IX70 microscope, a Coolsnap camera (Roper Scientific) and a $\times$ 100, 1.4 NA PlanApochromat objective. Images were computationally deconvolved using SoftWork (Applied Precision) and shown as two-dimensional projections.

Quantification of RPS4X expression

HeLa cells were transfected with esiRNA directed against *RPS4Y*, *RPS4X* or firefly luciferase (negative control). Cells were harvested 48 h after transfection. RNA was extracted using the RNeasy Mini Kit (Qiagen) including a DNaseI digest. cDNA was synthesized with SuperScript II reverse transcriptase (Invitrogen) using an oligo(dT) primer. *RPS4X* mRNA expression was measured by quantitative PCR using the Brilliant SYBR Green system and the Mx4000 Multiplex Quantitative PCR system (Stratagene). For *RPS4X* we used the primers (forward) 5'-CTGGATCTTTGACGTGGT-3' and (reverse) 5'-TTTCTCGGGGAAGAGAAAT-3', and for the reference gene *B2M* (GenBank accession number NM_004048) the primers were (forward) 5'-TGACTTTGTACAGCCCAAG-3' and (reverse) 5'-AGCAAGCAAGCAGAAATTGG-3'. Expression levels of *RPS4X* in cells transfected with *RPS4X*- and *RPS4Y*-specific esiRNA were normalized against the expression level of cells transfected with esiRNA targeting firefly luciferase. All experiments were performed in triplicate.

Received 13 July; accepted 1 November 2004; doi:10.1038/nature03159.

1. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
3. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
4. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
5. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
6. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
7. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
8. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
9. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
10. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
11. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
12. Yang, D. *et al.* Short RNA duplexes produced by hydrolysis with *Escherichia coli* RNase III mediate effective RNA interference in mammalian cells. *Proc. Natl Acad. Sci. USA* **99**, 9942–9947 (2002).
13. Zheng, L. *et al.* An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**, 135–140 (2004).

14. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
15. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
16. Calegari, F., Haubensack, W., Yang, D., Huttner, W. B. & Buchholz, F. Tissue-specific RNA interference in postimplantation mouse embryos with endoribonuclease-prepared short interfering RNA. *Proc. Natl Acad. Sci. USA* **99**, 14236–14240 (2002).
17. Kronke, J. *et al.* Alternative approaches for efficient inhibition of hepatitis C virus RNA replication by small interfering RNAs. *J. Virol.* **78**, 3436–3446 (2004).
18. Henschel, A., Buchholz, F. & Habermann, B. DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res.* **32**, W113–W120 (2004).
19. Liebel, U., Kindler, B. & Pepperkok, R. 'Harvester': a fast meta search engine of human protein resources. *Bioinformatics* **20**, 1962–1963 (2004).
20. Blangy, A. *et al.* Phosphorylation by p34cdc2 regulates spindle association of human Eg5, a kinesin-related motor essential for bipolar spindle formation *in vivo*. *Cell* **83**, 1159–1169 (1995).
21. Harborth, J., Elbashir, S. M., Bechert, K., Tuschl, T. & Weber, K. Identification of essential genes in cultured mammalian cells using small interfering RNAs. *J. Cell Sci.* **114**, 4557–4565 (2001).
22. Weis, K. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**, 441–451 (2003).
23. Biggins, S., Bhalla, N., Chang, A., Smith, D. L. & Murray, A. W. Genes involved in sister chromatid separation and segregation in the budding yeast *Saccharomyces cerevisiae*. *Genetics* **159**, 453–470 (2001).
24. Burns, C. G. *et al.* Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **22**, 801–815 (2002).
25. Prasanth, K. V., Sacco-Bubulya, P. A., Prasanth, S. G. & Spector, D. L. Sequential entry of components of the gene expression machinery into daughter nuclei. *Mol. Biol. Cell* **14**, 1043–1057 (2003).
26. Gruss, O. J. *et al.* Chromosome-induced microtubule assembly mediated by TPX2 is required for spindle formation in HeLa cells. *Nature Cell Biol.* **4**, 871–879 (2002).
27. Makarov, E. M. *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205–2208 (2002).
28. Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
29. Dorsett, Y. & Tuschl, T. siRNAs: applications in functional genomics and potential as therapeutics. *Nature Rev. Drug Discov.* **3**, 318–329 (2004).
30. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank F. Stewart, I. Baines, T. Hyman and M. Slabicki for critical reading and comments on the manuscript. We thank K. Weis for helpful discussions. We are grateful to M. Boutros for providing protein accession numbers and sequences of the cell viability screen in *Drosophila* cells. This study was supported by the Max Planck Society and by the EU-FP6 grant Mitochek. L.P. is supported by a postdoctoral fellowship from the HFSP.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.B. (buchholz@mpi-cbg.de). The sequences of KIAA1387 are deposited in GenBank under accession numbers AY825268 and AY825269.

Retinoblastoma promotes definitive erythropoiesis by repressing Id2 in fetal liver macrophages

Antonio Iavarone^{1,2,3}, Emerson R. King¹, Xu-Ming Dai⁶, Gustavo Leone⁵, E. Richard Stanley⁶ & Anna Lasorella^{1,2,4}

¹Institute for Cancer Genetics, Department of ²Pathology, ³Neurology and ⁴Pediatrics, College of Physicians and Surgeons of Columbia University, New York, New York 10032, USA

⁵Human Cancer Genetics Program, Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, Ohio 43210, USA

⁶Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

In mammals, the fetal liver is the first site of definitive erythropoiesis—the generation of mature, enucleated red cells. The functional unit for definitive erythropoiesis is the erythroblastic island, a multicellular structure composed of a central macrophage surrounded by erythroblasts at various stages of differen-

tiation^{1,2}. Targeted disruption of the retinoblastoma (*Rb*) tumour suppressor gene in the mouse leads to embryonic death caused by failure of erythroblasts to enucleate^{3–5}. The erythroid defect has been attributed to loss of *Rb* in cells that support erythropoiesis, but the identity of these cells is unknown⁶. Here we show that *Rb*-deficient embryos carry profound abnormalities of fetal liver macrophages that prevent physical interactions with erythroblasts. In contrast, wild-type macrophages bind *Rb*-deficient erythroblasts and lead them to terminal differentiation and enucleation. Loss of *Id2*, a helix–loop–helix protein that mediates the lethality of *Rb*-deficient embryos⁷, rescues the defects of *Rb*-deficient fetal liver macrophages. *Rb* promotes differentiation of macrophages by opposing the inhibitory functions of *Id2* on the transcription factor PU.1, a master regulator of macrophage differentiation. Thus, *Rb* has a cell autonomous function in fetal liver macrophages, and restrains *Id2* in these cells in order to implement definitive erythropoiesis.

The hallmark of the lethal anaemia of *Rb*-deficient embryos is the persistence of nucleated erythroid cells in the peripheral blood^{3–5}. However, when exposed to a wild-type stromal environment in chimaeric mice, *Rb*-deficient erythroblasts differentiate, enucleate and survive^{8,9}. The helix–loop–helix (HLH) factor *Id2* is an *Rb*-binding protein essential for lethality and phenotypic abnormalities of *Rb* mutant embryos, including the anaemia⁷. To identify gene targets of *Rb* regulated through its interaction with *Id2*, we performed microarray analysis with RNA isolated from mouse embryo fibroblasts (MEFs) derived from wild-type, *Rb*^{−/−} and *Id2*^{−/−}; *Rb*^{−/−} embryos at embryonic day E13.5. Compared to wild-type cells, a distinct group of genes whose expression is decreased in *Rb*^{−/−} MEFs and reversed and even upregulated in *Id2*^{−/−}; *Rb*^{−/−} cells belongs to the macrophage/dendritic cell lineage (Supplementary Fig. 1). Prominent among them is the gene coding for colony stimulating factor-1 receptor (*CSF-1R*), a receptor with a pivotal role in differentiation and survival of macrophages (Supplementary Fig. 2)^{10,11}. The first haematopoietic cells to emerge during mouse development are erythroblasts and primitive macrophages¹².

Because fetal liver macrophages (FLMs) are indispensable for definitive erythropoiesis¹³, we examined the role of *Rb* in development of FLMs. First, we asked whether *Rb*-deficient embryos have reduced expression of *CSF-1R* in the fetal liver. Western blot analysis of *CSF-1R* from E13.5 fetal livers shows that *CSF-1R* is strongly expressed in lysates derived from wild-type livers but is almost undetectable in *Rb*^{−/−} livers. Deletion of *Id2* in the *Rb*-deficient background restores expression of *CSF-1R* with a clear *Id2*-dose dependency (Fig. 1a). Next, we tested whether *Rb*-deficient embryos carry *Id2*-dependent abnormalities of FLMs that might explain the erythropoietic failure (Fig. 1b). Staining for the mature macrophage marker F4/80 shows that *Rb*-deficient fetal livers have a dramatic decrease in the number of FLMs (Fig. 1c). The scarce F4/80-positive macrophages show weak staining, are generally round and small, and lack the extensive cytoplasmic projections of normal FLMs, suggesting an immature morphology (Fig. 1d). F4/80-positive FLMs derived from *Id2*^{−/−}; *Rb*^{−/−} livers at E13.5 and E15.5, an age when the *Rb*-deficient embryos are in an advanced state of resorption, are indistinguishable from wild-type embryos (Figs 1c, d, and 2b, upper panel). The unrelated megakaryocyte marker von Willebrand factor (vWF) and, more importantly, ER-MP20 and ER-MP12, two antigens expressed by immature macrophage precursors¹⁴, are similar in wild-type, *Rb*^{−/−} and *Id2*^{−/−}; *Rb*^{−/−} fetal livers (Fig. 1e–g). These results suggest that unrestricted *Id2* activity in *Rb*-deficient FLMs impairs terminal differentiation but not commitment to the macrophage lineage.

Double immunostaining of E13.5 and E15.5 wild-type and *Id2*^{−/−}; *Rb*^{−/−} fetal livers for F4/80 and the erythroid antigen TER119 shows that the entire perimeter of the macrophage membranes adhere to erythroid cells (Fig. 2a, b, lower panels). Conversely, the small and

immature *Rb*^{−/−} FLMs fail to establish significant contacts with erythroblasts *in vivo* (Fig. 2a). Consistent with a role for *Rb* and *Id2* in regulating development of FLMs, double immunostaining of wild-type FLMs revealed that the nuclei of F4/80-positive FLMs stain positively for *Id2* and *Rb* (Fig. 2c, d). To investigate whether endogenous *Id2* and *Rb* interact in macrophages, we cultured primary mouse macrophages in the presence of CSF-1. First, we performed quantitative western blot analysis⁷ and determined that the molar amount of hypophosphorylated *Rb* (p*Rb*) in macrophages exceeds that of *Id2* by a factor of approximately ten (p*Rb*: 33.0 pmol per 10⁶ cells; *Id2*: 3.5 pmol per 10⁶ cells). Then, through immunoprecipitation-western blot experiments, we show that endogenous *Id2* interacts efficiently with p*Rb* in macrophages (Fig. 2e).

To find out whether *Rb* is essential for the ability of FLMs to bind erythroblasts, we developed a macrophage-erythroblast co-culture assay *in vitro* (see Supplementary Information). Livers from E13.5 embryos are partially digested with collagenase and FLMs are

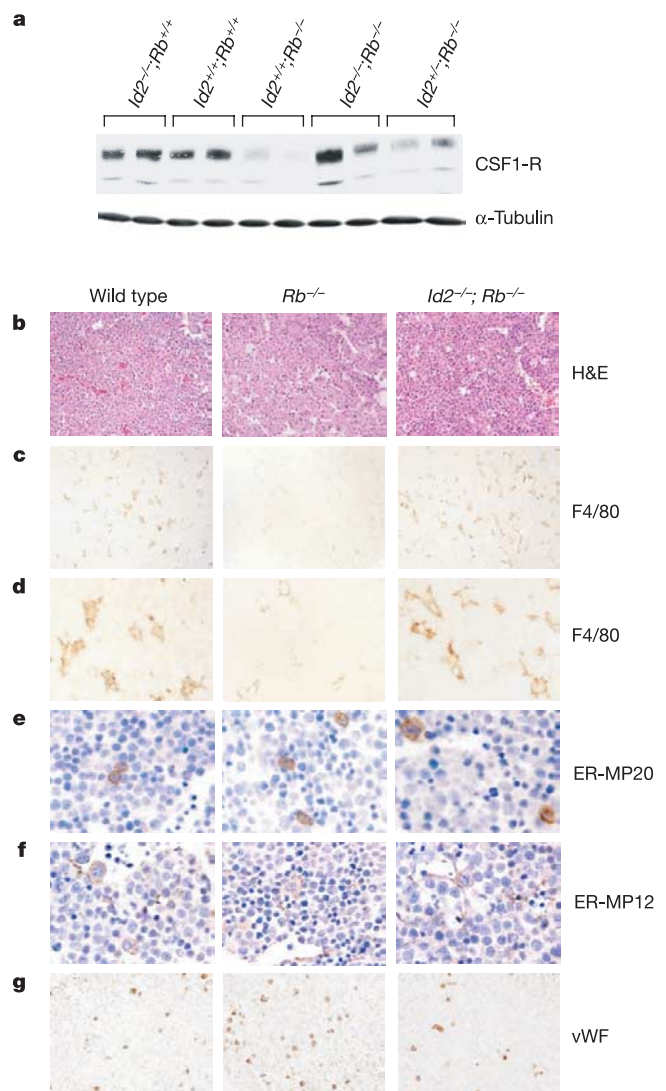


Figure 1 Abnormal development of FLMs in *Rb*^{−/−} embryos and rescue by loss of *Id2*. **a**, Western blot analysis for *CSF-1R* and α -tubulin from E13.5 livers of the indicated genotypes (two embryos per genotype). Liver sections from E13.5 wild-type, *Rb*^{−/−} and *Id2*^{−/−}; *Rb*^{−/−} embryos were stained with haematoxylin and eosin (H&E, **b**), and immunostained with F4/80 (**c**, **d**), ER-MP20 (**e**), ER-MP12 (**f**) and vWF (**g**). Original magnification: $\times 20$ (**b**, **c**, **g**); $\times 60$ (**d**–**f**).

selectively allowed to adhere to a substrate for a short-term incubation. Confocal microscopy images of these cultures show numerous clusters composed of TER119-positive erythroblasts attached to large substrate-adherent F4/80-positive macrophages ('native' islands, Fig. 3a). Treatment of the clusters with a solution lacking calcium and magnesium removes the erythroblasts, leaving intact the adherent macrophages ('stripped' islands, Fig. 3b). Under our experimental conditions, 95% of adherent cells stain positively for F4/80, indicating that they are FLMs (data not shown).

Next, we challenged the 'naked' macrophages with heterologous, freshly isolated erythroblasts from E14.5 fetal livers ('reconstituted' islands, Fig. 3c). Erythroblasts do not bind to 3T3 fibroblasts, indicating that the interaction between FLMs and erythroblasts is specific (data not shown). We find a 5.4-fold reduction of adherent FLMs from *Rb*-deficient livers versus a 1.5-fold decrease of FLMs from *Id2*^{-/-}; *Rb*^{-/-} livers (wild type, 4,505 ± 449, *n* = 7; *Rb*^{-/-}, 831 ± 339, *n* = 4; *Id2*^{-/-}; *Rb*^{-/-}, 2,877 ± 601, mean ± s.e.m., *n* = 6). Moreover, 8.3 ± 0.7 erythroblasts are attached to a native wild-type FLM, whereas only 0.1 ± 0.01 erythroblasts are recovered in association with an *Rb*-deficient FLM (*P* < 0.0001, Fig. 3a–d). The number of erythroblasts bound to wild-type FLMs in the reconstituted erythroblastic islands (8.7 ± 1.4 erythroblasts per macrophage, *n* = 4) is similar to the number of erythroblasts in the native clusters (Fig. 3a, c–e). *Rb*^{-/-} FLMs fail to adhere to

normal erythroblasts in the reconstitution experiment (0.4 ± 0.2 erythroblasts per macrophage, *n* = 6, *P* < 0.0001, Fig. 3c–e). In the native and reconstitution experiments, *Id2*^{-/-}; *Rb*^{-/-} FLMs display an almost normal ability to form clusters with erythroblasts (6.7 ± 3.6 erythroblasts per macrophage, *n* = 6, and 6.3 ± 2.7 erythroblasts per macrophage, *n* = 4, respectively, Fig. 3). Our results indicate that *Rb*-deficient FLMs are intrinsically defective in binding erythroblasts. They also show that the phenotypic abnormalities of *Rb*-deficient FLMs are effectively rescued *in vivo* and *in vitro* by mutation of *Id2*.

We sought to determine whether *Rb*-deficient erythroblasts retain the ability to bind normal FLMs. To obtain the number of cells required for these experiments, we expanded erythroid progenitors from the fetal liver of wild-type and *Rb*^{-/-} embryos. In the presence of stem cell factor, dexamethasone, insulin-like growth factor-1 and erythropoietin, the proliferation of *Rb*-deficient erythroid progenitors is indistinguishable from wild-type cells (Fig. 3f). When cultured for 20 min with wild-type FLMs that had been stripped of endogenous erythroblasts, *Rb*-deficient erythroblasts bind avidly to FLMs (Fig. 3g). We reasoned that if the deficient contacts provided by FLMs are responsible for the failure of *Rb*^{-/-} erythroblasts to enucleate, the assembly of erythroblastic islands composed of wild-type FLMs and *Rb*^{-/-} erythroid cells *in vitro* should rescue the enucleation defect. We therefore developed culture conditions that allow survival and FLM-mediated enucleation of bound erythroblasts. We first allow the binding of erythroblasts to heterologous, stripped FLMs and then culture the reconstituted erythroblastic islands in the presence of erythropoietin and CSF-1. In the absence of FLMs, wild-type and *Rb*^{-/-} erythroid cells cultured in the same medium or in FLM-conditioned medium are at the proerythroblast and basophilic normoblast stage with large nucleus and weak TER119 staining (Fig. 3h). Conversely, following binding to FLMs, a significant fraction of erythroid cells become strongly positive for haemoglobin, TER119 and enucleate (Fig. 3i, j and data not shown). The percentage of FLM-induced enucleation of wild-type and *Rb*^{-/-} erythroid cells is similar (wild type: 38%; *Rb*^{-/-}: 37%). These results indicate that physical interaction with wild-type FLMs rescues the enucleation defect of *Rb*-deficient erythroblasts.

The CSF-1R is critical for the development of most tissue macrophages^{11,15}. Therefore, we asked whether the deregulated *Id2* activity resulting from loss of *Rb* would be sufficient to affect expression of CSF-1R. Treatment of the myeloid cell line U937 with the macrophage differentiating agents retinoic acid and 1,25-dihydroxyvitamin D3 induces the expression of CSF-1R¹⁶. Infection of U937 with an *Id2*-expressing adenovirus abolishes the basal expression of CSF-1R, and leads to an 80% decrease of the retinoic acid/1,25-dihydroxyvitamin D3-mediated induction of CSF-1R (Fig. 4a). The Ets-related protein PU.1 is a transcription factor that directs expression of a number of macrophage-specific genes, including *CSF-1R*^{17,18}. As *Id* proteins have been shown to bind and antagonize the activity of transcription factors of the Ets family¹⁹, it seems possible that *Id2* might inhibit PU.1. First, we determine that untagged and FLAG-tagged PU.1 bind GST-*Id2* *in vitro* with equal affinity. A GST-*Id2* mutant lacking the critical HLH domain fails to bind PU.1 (Supplementary Fig. 3). After co-transfecting cells with plasmids encoding *Id2* and Flag-PU.1, we demonstrate co-precipitation of the two proteins with antibodies against either Flag or *Id2* (Fig. 4b).

Transcription of *CSF-1R* is regulated by a highly conserved enhancer element in the first intron of the gene, referred to as *Fms* intronic regulatory element (*FIRE*)²⁰. During macrophage differentiation, PU.1 binds *in vivo* to two adjacent PU-consensus sites in the *FIRE* region of *CSF-1R*²¹. To determine whether the interaction of *Id2* with PU.1 prevents binding of PU.1 to important regulatory elements in the *CSF-1R-FIRE* region, we performed electrophoretic mobility shift (EMSA) and chromatin immuno-

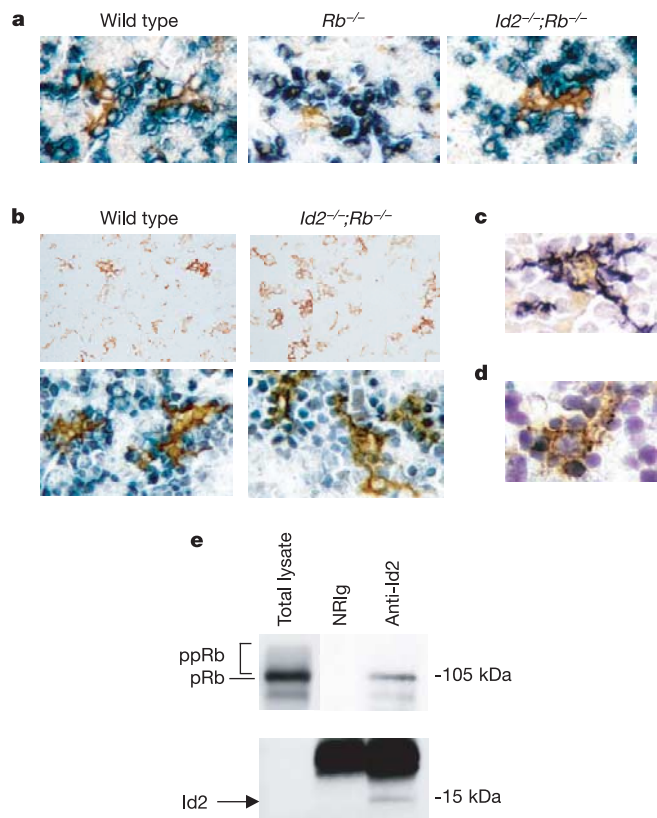


Figure 2 Deregulated *Id2* causes defective macrophage-erythroblast interactions in the *Rb*-deficient fetal liver. **a**, E13.5 wild-type, *Rb*^{-/-} and *Id2*^{-/-}; *Rb*^{-/-} fetal livers were stained for F4/80 (brown) and TER119 (blue). **b**, E15.5 wild-type and *Id2*^{-/-}; *Rb*^{-/-} fetal livers were stained for F4/80 (upper panels) and F4/80 plus TER119 (lower panels). Wild-type E15.5 livers were stained for F4/80 (blue) and *Id2* (brown/yellow) (**c**), or for F4/80 (brown) and *Rb* (purple) (**d**). Original magnification: × 40 (**b**, upper panels); × 100 (**a**, **b**-lower panels, **c**, **d**). **e**, Immunoprecipitates from primary mouse macrophages with anti-*Id2* antibodies or normal rabbit immunoglobulin (NRIg) were analysed by western blot for *Rb* (upper panel) and *Id2* (lower panel). pRb, hypophosphorylated *Rb*, ppRb, hyperphosphorylated *Rb*.

precipitation (ChIP) assays. Expression of PU.1 leads to the formation of a specific PU.1 complex that, in EMSA assays, binds to an oligonucleotide containing the two *PU*-DNA binding elements from the mouse *CSF-1R-FIRE* region. Co-expression of Id2 results in a dose-dependent inhibition of the binding of PU.1 to DNA (Supplementary Fig. 4). In ChIP assays, Flag-PU.1 readily binds *CSF-1R-FIRE* *in vivo* and Id2 inhibits the binding. Interestingly, the Id2 inhibition is stoichiometrically rescued by co-expression of a constitutively active, unphosphorylatable Rb protein (PSM-Rb), which retains its ability to bind Id2 (Fig. 4c; see also Supplementary Fig. 5, ref. 28 and data not shown).

To determine the consequences of the *Rb* mutation for the binding of PU.1 to the *CSF-1R-FIRE* regulatory region in primary cells, we cultured macrophage progenitors from the fetal liver of wild-type and *Rb*^{-/-} embryos in the presence of IL-3, IL-6, Flt3 ligand and GM-CSF. Under these experimental conditions, seven-day-old cultures derived from wild-type fetal livers contain approximately equal numbers of adherent, differentiated macrophages and floating, more immature monocytes. *Rb*-deficient myeloid cultures lack mature and adherent macrophages and show strongly reduced expression of the late macrophage differentiation markers and PU.1 targets, namely CSF-1R, scavenger receptor A (SR-A) and CD11b (Fig. 4d). When the same cultures were used to investigate by ChIP the binding of endogenous PU.1 to *CSF-1R-FIRE*, we found that PU.1 is bound to the *CSF-1R-FIRE* element in wild-type but not *Rb*-deficient cells (Fig. 4e). We finally asked whether expression of Id2 would affect the ability of PU.1 to activate transcription through the *CSF-1R-FIRE* enhancer element. With a reporter construct

including *FIRE* upstream of the *CSF-1R* core promoter, we observe a dose-dependent reduction of PU.1-mediated transcription by Id2. Again, the Id2 inhibition is fully reversed by an excess of active Rb (Fig. 4f). Id2 also inhibits luciferase activity driven by transcriptional activation of three multimerized *PU* boxes by PU.1, thus acting as a general inhibitor of PU.1-mediated transcription (Fig. 4g). It has been reported that Rb binds directly to PU.1²². Therefore, we asked whether, regardless of Id2, Rb facilitates activation of CSF-1R transcription by PU.1. Expression of PSM-Rb in the absence of Id2 does not stimulate PU.1-dependent transcription from the *CSF-1R-FIRE* or the multimerized *PU* boxes, indicating that Rb operates as a specific inhibitor of Id2 to enhance the expression of PU.1 target genes (Fig. 4h and data not shown). Taken together, these data suggest that the ratio of Id2 to active Rb is a key factor for DNA binding and transcriptional activation of *CSF-1R* by PU.1.

We have shown that the Id2-dependent phenotype of *Rb*-deficient FLMs is a crucial component of the non-cell autonomous anaemia of *Rb* mutant embryos. In an *in vitro* culture assay, *Rb*-deficient erythroblasts display a failure to enucleate²³. However, when provided with physical contacts with wild-type FLMs, *Rb*-deficient erythroid cells reconstitute normal erythroblastic islands and undergo differentiation and enucleation. The abnormal development of FLMs in *Rb*-deficient embryos is consistent with an earlier study showing that antisense oligonucleotides against Rb inhibit differentiation of bone marrow progenitors along the monocyte lineage²⁴. In addition to mutation of *Id2* and *E2F* transcription factors^{7,25,26}, rescue of the erythropoietic defect of

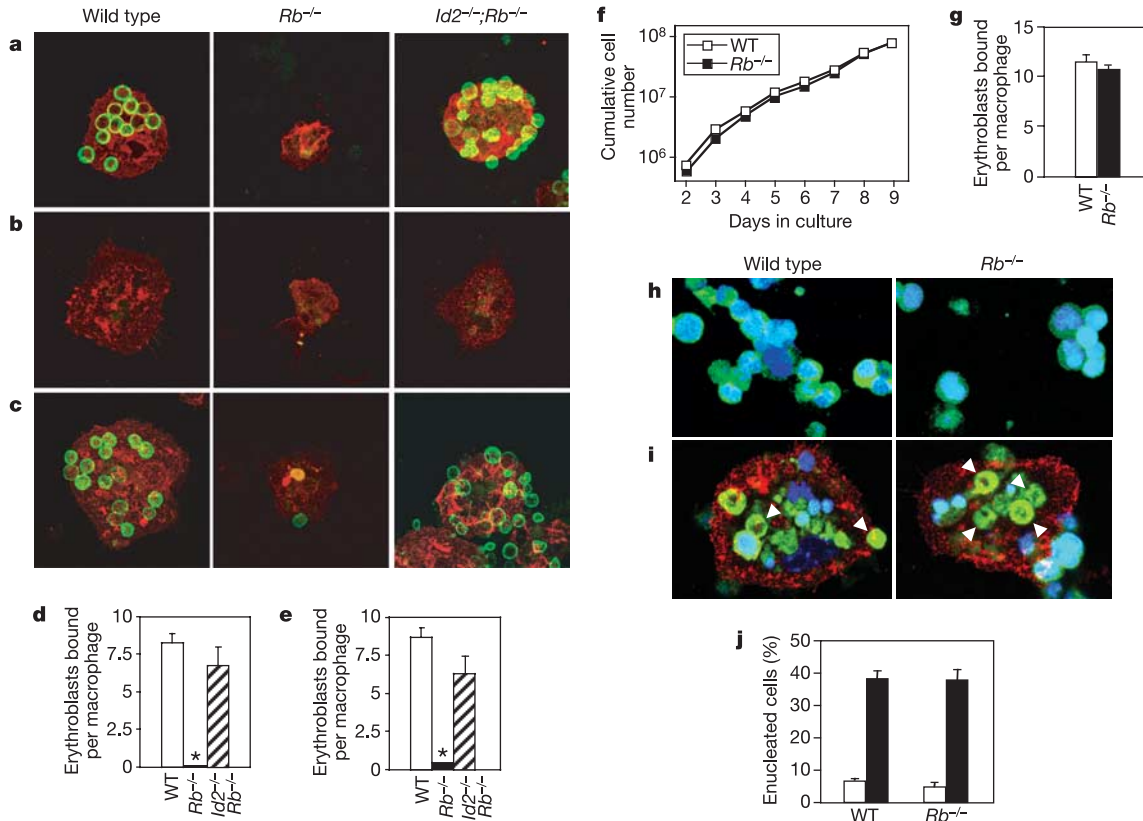


Figure 3 *Rb*-deficient erythroblasts but not *Rb*-deficient FLMs reconstitute functional erythroblastic islands. **a**, Native islands were stained for F4/80 (red) and TER119 (green). Stripped FLMs (**b**) were reconstituted with wild-type erythroid cells (**c**). Erythroblasts bound per macrophage in native (**d**) and reconstituted (**e**) islands. **P* < 0.0001. **f**, Proliferation of wild-type (*n* = 10) and *Rb*^{-/-} (*n* = 4) erythroblasts. **g**, FLMs

reconstituted with wild-type (white bar) and *Rb*^{-/-} (black bar) erythroblasts (*n* = 3). Erythroblasts grown in the absence (**h**) or in association with wild-type FLMs (**i**) were stained for F4/80 (red), TER119 (green) and DAPI (blue). Arrowheads, enucleated cells. **j**, Percentage of enucleated cells in the absence (white bars) or in association with wild-type FLMs (black bars) (*n* = 4).

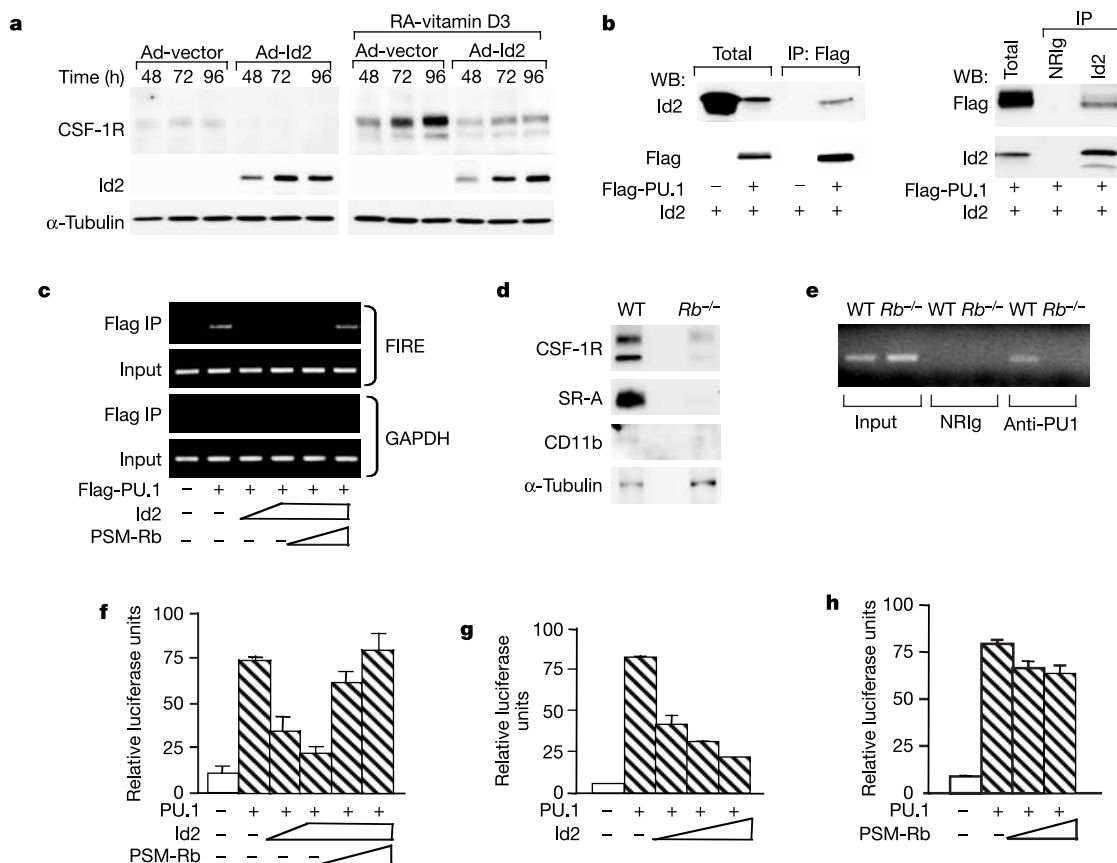


Figure 4 Id2 and Rb regulate transcriptional activation of *CSF1R* by PU.1. **a**, Western blot analysis for CSF-1R, Id2 and α -tubulin from U937 infected with Ad-GFP or Ad-Id2 untreated or treated with retinoic acid plus vitamin D3. **b**, Lysates from transfected COS-1 were analysed directly (total) or immunoprecipitated with antibodies against Flag (left panel) or Id2 (right panel). NRlg, normal rabbit immunoglobulins. **c**, Anti-Flag

immunoprecipitates from transfected NIH 3T3 were analysed by PCR for *CSF-1R-FIRE* and *GAPDH*. Fetal liver-derived macrophages were analysed for **(d)** CSF-1R, SR-A, CD11b and α -tubulin by western blot and **(e)** ChIP to test the binding of PU.1 to *CSF-1R-FIRE*. **f–h**, Transcription from a *CSF-1R-FIRE*-luciferase plasmid (**f**), or a multimerized *PU*-box-luciferase (**g**).

Rb-deficient embryos has been achieved by restoring *Rb* function in extra-embryonic tissues²⁷. It will be interesting to test whether distinct genetic systems converge on restoring functionally competent FLMs to implement the rescue of the anaemia. Our finding of the molecular cross talk between Rb, Id2 and PU.1 in the transcriptional control of *CSF-1R* provides a mechanistic model for the rescue of the erythropoietic defect of *Rb*-deficient embryos lacking *Id2*. Thus, inhibition of Id2 is a general mechanism by which Rb promotes the activity of tissue-specific transcription factors and differentiation^{7,28}.

Methods

Mouse strains and genotyping

Rb^{+/-} and *Id2*^{+/-} mice, both in a mixed C57B6/129 Sv background, were crossed to generate *Id2*^{+/-}; *Rb*^{+/-} mice that were then crossed with each other to generate the embryos described in this study. Mice were genotyped by PCR as described⁷ and maintained in compliance with IACUC guidelines.

Histological analysis and immunohistochemistry

Freshly collected livers from dated pregnancies were fixed in 10% neutral formalin, embedded in paraffin and cut as 5- μ m serial sections, which were stained with haematoxylin and eosin or processed for immunohistochemistry as described⁷. The antibodies used in immunohistochemical analysis were: F4/80 (clone A3-1, Caltag Laboratories), TER119 (Caltag Laboratories), ER-MP12 (BMA Biomedicals), ER-MP20 (BMA Biomedicals), vWF (Dako), Rb (G3-245, BD Pharmingen), Id2 (C-20, Santa Cruz). The avidin-biotin complex (ABC) technique was used for primary antibody detection.

Cell culture

U937 cells were infected with an adenoviral vector expressing Id2 or Ad-GFP vector at MOI of 200. Cells were collected at the indicated times and processed for western blot. For

differentiation experiments, infected U937 cells were exposed to 1 μ M all-trans retinoic acid (Sigma) plus 0.1 μ M 1,25-dihydroxyvitamin D3 (Sigma). Methods for isolation, stripping, reconstitution and functional analysis of FLM-erythroblast clusters are provided as Supplementary Methods.

Protein analysis

Immunoblotting and immunoprecipitation were performed as described⁷ with primary antibodies against Flag (M2, Sigma), Id2 (C-20, Santa Cruz), PU.1 (T-21, Santa Cruz), CSF-1R (C-20, Santa Cruz), Rb (XZ55, BD Pharmingen), SR-A (Serotec) and CD11b (DakoCytomation). The *in vitro* binding assays were performed as described²⁹. To analyse Id2 immunoprecipitates in primary mouse macrophages, bone-marrow-derived macrophages were prepared as described previously³⁰ and cultured in supplemented α -modified minimal essential medium (α -MEM, Invitrogen) for five days in the presence of 1,000 U ml⁻¹ human recombinant CSF-1. Cell lysates were immunoprecipitated with polyclonal antibodies against Id2 (C-20, Santa Cruz) and analysed by western blot for Rb (G3-245, BD Pharmingen) and Id2. To measure the endogenous amounts of Id2 and hypophosphorylated Rb expressed by primary mouse macrophages, quantitative western blots were performed as described⁷.

EMSA, ChIP and promoter analysis

Gel retardation assays were performed with ³²P-labelled oligonucleotides including two *PU* boxes from the mouse *CSF-1R-FIRE* region (top strand, 5'-GTTCTCACTTCCCCCCTCCCCCTAT). DNA-protein complexes were formed at room temperature for 20 min using cellular extracts from COS cells that had been transfected with plasmids expressing Flag-PU.1 and Flag-Id2. Protein-DNA complexes were analysed on nondenaturing 4% polyacrylamide gel in 1 $\times$ Tris-borate-EDTA buffer and visualized by autoradiography. ChIP assays were performed using a modification of a published protocol⁷. One day after seeding, NIH 3T3 cells were transfected with the plasmid pGL6.7 Δ fms (ref. 20) containing the *CSF-1R* genomic locus plus CMV expression vectors for Flag-PU.1, Id2 and PSM-Rb using Lipofectamine 2000 (Invitrogen). After immunoprecipitation with the M2 anti-Flag antibody, the recovered DNA was analysed by PCR with primers flanking the *FIRE* element (5'-CAGTGTGTGAGAAGGACAATGGC-3' and 5'-GCTTGAAGGCGAAGAGAGAACTC-3') and the *GAPDH* gene (5'-TGGTATCGTGAAGGACTCATGAC-3' and

5'-ATGCCAGTGAAGCTTCCCGTTTCAGC-3'). Primary cultures of fetal-liver-derived macrophage progenitors were generated from E12.5 *Rb*^{+/+} and *Rb*^{-/-} embryos and cultured in α -MEM supplemented with 10% FBS plus SCF (100 ng ml⁻¹, R&D Systems), IL-3 (10% medium conditioned by WEHI-2 cells, a gift from N.L. Warner), recombinant Flt3 ligand (10 ng ml⁻¹, R&D Systems), recombinant IL-6 (5 ng ml⁻¹, R&D Systems) and GM-CSF (2% medium conditioned by MELrmGMCSF cells, a gift from N.A. Nicola). After seven days of culture, cells were used for western blot analysis and ChIP assays. Immunoprecipitations were done with normal rabbit immunoglobulins and polyclonal antibodies against PU.1 (T-21, Santa Cruz). For promoter-luciferase assays, NIH 3T3 cells were transfected with a *CSF-1R-FIRE*-luciferase plasmid (Δ *fms-FIRE*) containing the 300-bp *FIRE* sequence upstream of the *CSF-1R* minimal promoter²⁰ or a plasmid expressing luciferase from three multimerized *PU*-boxes. Transfected cells were lysed and luciferase assays were carried out using the Promega luciferase assay kit and a Berthold luminometer. Normalized values are reported as the mean $\pm$ s.d. from duplicate transfections.

Received 23 August; accepted 25 September 2004; doi:10.1038/nature03068.

- Sasaki, K. & Iwatsuki, H. Origin and fate of the central macrophages of erythroblastic islands in the fetal and neonatal mouse liver. *Microsc. Res. Tech.* **39**, 398–405 (1997).
- Naito, M., Hasegawa, G. & Takahashi, K. Development, differentiation, and maturation of Kupffer cells. *Microsc. Res. Tech.* **39**, 350–364 (1997).
- Clarke, A. R. et al. Requirement for a functional *Rb-1* gene in murine development. *Nature* **359**, 328–330 (1992).
- Lee, E. Y.-H. P. et al. Mice deficient for *Rb* are nonviable and show defects in neurogenesis and hematopoiesis. *Nature* **359**, 288–294 (1992).
- Jacks, T. et al. Effects of an *Rb* mutation in the mouse. *Nature* **359**, 295–300 (1992).
- Vooijs, M. & Berns, A. Developmental defects and tumor predisposition in *Rb* mutant mice. *Oncogene* **18**, 5293–5303 (1999).
- Lasorella, A., Noseda, M., Beyna, M., Yokota, Y. & Iavarone, A. Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins. *Nature* **407**, 592–598 (2000).
- Maandag, E. C. et al. Developmental rescue of an embryonic-lethal mutation in the retinoblastoma gene in chimeric mice. *EMBO J.* **13**, 4260–4268 (1994).
- Williams, B. O. et al. Extensive contribution of *Rb*-deficient cells to adult chimeric mice with limited histopathological consequences. *EMBO J.* **13**, 4251–4259 (1994).
- Pixley, F. J. & Stanley, E. R. CSF-1 regulation of the wandering macrophage: Complexity in action. *Trends Cell Biol.* **114**, 628–638 (2004).
- Cecchini, M. G. et al. Role of colony stimulating factor-1 in the establishment and regulation of tissue macrophages during postnatal development of the mouse. *Development* **120**, 1357–1372 (1994).
- Moore, M. A. & Metcalf, D. Ontogeny of the haemopoietic system: yolk sac origin of *in vivo* and *in vitro* forming cells in the developing mouse embryo. *Br. J. Haematol.* **18**, 279–296 (1970).
- Kawane, K. et al. Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science* **292**, 1546–1549 (2001).
- Morioka, Y., Naito, M., Sato, T. & Takahashi, K. Immunophenotypic and ultrastructural heterogeneity of macrophage differentiation in bone marrow and fetal hematopoiesis of mouse *in vitro* and *in vivo*. *J. Leukoc. Biol.* **55**, 642–651 (1994).
- Dai, X. M. et al. Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood* **99**, 111–120 (2002).
- Nakajima, H. et al. All-trans and 9-cis retinoic acid enhance 1,25-dihydroxyvitamin D₃-induced monocytic differentiation of U937 cells. *Leuk. Res.* **20**, 665–676 (1996).
- DeKoter, R. P., Walsh, J. C. & Singh, H. PU.1 regulates both cytokine-dependent proliferation and differentiation of granulocyte/macrophage progenitors. *EMBO J.* **17**, 4456–4468 (1998).
- Anderson, K. L. et al. Myeloid development is selectively disrupted in PU.1 null mice. *Blood* **91**, 3702–3710 (1998).
- Yates, P. R., Atherton, G. T., Deed, R. W., Norton, J. D. & Sharrocks, A. D. Id helix-loop-helix proteins inhibit nucleoprotein complex formation by the TCF ETS-domain transcription factors. *EMBO J.* **18**, 968–976 (1999).
- Himes, S. R. et al. A highly conserved c-fms gene intronic element controls macrophage-specific and regulated expression. *J. Leukoc. Biol.* **70**, 812–820 (2001).
- Tagoh, H. et al. Transcription factor complex formation and chromatin fine structure alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes Dev.* **16**, 1721–1737 (2002).
- Rekhtman, N. et al. PU.1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation. *Mol. Cell. Biol.* **23**, 7460–7474 (2003).
- Clark, A. J., Doyle, K. M. & Humbert, P. O. Cell-intrinsic requirement for pRB in erythropoiesis. *Blood* **104**, 1324–1326 (2004).
- Bergh, G., Ehinger, M., Olsson, I., Jacobsen, S. E. & Gullberg, U. Involvement of the retinoblastoma protein in monocytic and neutrophilic lineage commitment of human bone marrow progenitor cells. *Blood* **94**, 1971–1978 (1999).
- Tsai, K. Y. et al. Mutation of E2f-1 suppresses apoptosis and inappropriate S phase entry and extends survival of *Rb*-deficient mouse embryos. *Mol. Cell* **2**, 293–304 (1998).
- Ziebold, U., Reza, T., Caron, A. & Lees, J. A. E2F3 contributes both to the inappropriate proliferation and to the apoptosis arising in *Rb* mutant embryos. *Genes Dev.* **15**, 386–391 (2001).
- Wu, L. et al. Extra-embryonic function of *Rb* is essential for embryonic development and viability. *Nature* **421**, 942–947 (2003).
- Toma, J. G., El-Bizri, H., Barnabe-Heider, F., Aloyz, R. & Miller, F. D. Evidence that helix-loop-helix proteins collaborate with retinoblastoma tumor suppressor protein to regulate cortical neurogenesis. *J. Neurosci.* **20**, 7648–7656 (2000).
- Iavarone, A., Garg, P., Lasorella, A., Hsu, J. & Israel, M. A. The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. *Genes Dev.* **8**, 1270–1284 (1994).
- Stanley, E. R. Murine bone marrow-derived macrophages. *Methods Mol. Biol.* **75**, 301–304 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Tycko for discussions, Y. Yokota and T. Jacks for providing us with *Id2*^{+/+} and *Rb*^{+/+} mice, T. Ludwig for support in the cultures of FLMs and erythroblasts, and R. Russell for histological consultation. We thank D. Hume for *CSF-1R* promoter plasmids, T. Graf for *PU.1* expression plasmids, E. Knudsen for the *PSM-Rb* plasmid and R. Wu for help with ER-MP12 and ER-MP20 immunostaining. This work was supported by grants from NIH to A.I., A.L., G.L. and E.R.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.I. (ai2102@columbia.edu).

Hedgehog signalling activity of Smoothed requires phosphorylation by protein kinase A and casein kinase I

Jianhang Jia^{1*}, Chao Tong^{1*}, Bing Wang¹, Liping Luo¹ & Jin Jiang^{1,2}

¹Center for Developmental Biology and ²Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

* These authors contributed equally to this work

The Hedgehog (Hh) family of secreted proteins governs cell growth and patterning in animal development¹. The Hh signal is transduced by the seven-transmembrane protein Smoothed (Smo); however, the mechanism by which Smo is regulated remains largely unknown. Here we show that protein kinase A (PKA) and casein kinase I (CKI) regulate Smo cell-surface accumulation and activity in response to Hh. Blocking PKA or CKI activity in the *Drosophila* wing disc prevents Hh-induced Smo accumulation and attenuates pathway activity, whereas increasing PKA activity promotes Smo accumulation and pathway activation. We show that PKA and CKI phosphorylate Smo at several sites, and that phosphorylation-deficient forms of Smo fail to accumulate on the cell surface and are unable to transduce the Hh signal. Conversely, phosphorylation-mimicking Smo variants show constitutive cell-surface expression and signalling activity. Furthermore, we find that the levels of Smo cell-surface expression and activity correlate with its levels of phosphorylation. Our data indicate that Hh induces progressive Smo phosphorylation by PKA and CKI, leading to elevation of Smo cell-surface levels and signalling activity.

In many developmental processes, Hh acts as a morphogen to specify different developmental outcomes as a function of its concentration². In the developing *Drosophila* wing, posterior (P) compartment cells express and secrete Hh proteins which move into the anterior (A) compartment and form a local activity gradient. Low levels of Hh suffice to activate Hh target genes such as *decapentaplegic* (*dpp*), whereas high levels of Hh are required to activate other genes including *patched* (*ptc*), *collier* (*col*), and *engrailed* (*en*) (refs 3, 4).

The reception system for Hh consists of two transmembrane proteins, Patched (Ptc) and Smo (ref. 1). Ptc inhibits Smo, which is alleviated by Hh binding to Ptc. The activated Smo elicits an intracellular signalling cascade culminating in activation of the latent transcription factor Cubitus interruptus (Ci).

PKA was identified as an inhibitory component of the Hh pathway^{5–8}. PKA fulfills its negative role by phosphorylating full-length Ci (Ci155) at several Ser/Thr residues, priming it for further

phosphorylation by glycogen synthase kinase 3 and CKI (refs 9–12). Hyperphosphorylation of Ci155 targets it for proteolytic processing to generate the repressor form (Ci75) (ref. 13). Consistent with this, overexpressing a constitutively active form of PKA catalytic subunit (PKAc), mC*, blocks Ci155 accumulation and Hh target gene expression⁹. Surprisingly, we found that a modest increase of mC* resulted in enhanced Hh pathway activity. Mis-expression of a weak line of *UAS-mC** using a wing-specific Gal4 driver, *MS1096* (ref. 9), resulted in anterior expansion of both *ptc* and *en* expression

domains (compare Fig. 1e with b; data not shown). High levels of Hh induce maturation of Ci155 into a labile, hyperactive form so that A-compartment cells abutting the A/P border show low levels of Ci155 (Fig. 1a)¹⁴. We found that the 'low Ci155' domain expands concomitantly with the anterior *en* expression domain (compare Fig. 1d with a), indicating that increasing PKA activity elevates Hh signalling activity, and converts Hh responses from low to high in more anterior cells.

Expressing mC* using *hh-Gal4* did not alter Hh responses in A-compartment cells, whereas expressing mC* using *dpp-Gal4* resulted in similar enhancement of Hh signalling activity (data not shown). Moreover, expressing mC* with *apterous* (*ap*)-*Gal4*, a dorsal (D) compartment-specific Gal4 driver, caused anterior expansion of Hh target genes such as *col* only in D-compartment cells (Fig. 1g, h). These observations indicate that mC* exerts a positive, cell-autonomous influence on Hh signalling in responding cells.

Smo is accumulated in P-compartment cells, as well as A-compartment cells near the A/P boundary¹⁵. The elevation of Smo is more pronounced in A-compartment cells abutting the A/P border where *en* is expressed¹⁵ (Fig. 1c). We found that mC* induced anterior expansion of the 'high Smo' domain (Fig. 1f, i), indicating that PKA may promote pathway activation by regulating Smo accumulation.

We asked whether PKA activity is required for Hh-induced Smo accumulation and high-threshold Hh responses. To block PKA activity, we overexpressed a mutant PKA regulator subunit, R*, which binds cyclic AMP poorly and therefore keeps PKAc in an inactive holoenzyme¹⁶. In wing discs expressing R* with *ap-Gal4*, Hh-induced Smo accumulation and *en* expression were blocked in D-compartment cells (Fig. 1k, l). In addition, *col* expression at the A/P border was attenuated (Fig. 1j).

We examined Smo in wing discs carrying mitotic clones mutant for *DC0*, which encodes the main isoform of PKAc in *Drosophila*¹⁷, and found that Smo accumulation was attenuated in early-induced large *DC0*⁻ clones (Fig. 1m–o). Moreover, A-compartment *DC0*⁻ cells abutting the A/P border showed reduced levels of *en* expression (data not shown). Overall, the effect of *DC0* mutation on Smo accumulation and *en* expression is less severe than that caused by expressing R*, probably owing to compensation by other PKAc isoforms. This indicates that low levels of constitutive PKA activity are required for Smo accumulation and high-threshold Hh responses.

Hh induces Smo phosphorylation¹⁵; however, the relevant kinase(s) remains elusive. The Smo C-terminal tail contains five predicted PKA sites^{18,19}, including Ser 667, Ser 687 and Ser 740 that fit the consensus: RR/KXS (ref. 20) (Fig. 2a). (Here bold, italic residue indicates the phospho-receptor site.) Interestingly, each PKA site precedes two predicted CKI sites of the consensus: (p)SXXSXXS (ref. 21). In addition, these predicted phosphorylation clusters are conserved in *Anopheles gambiae* Smo (Fig. 2b). To determine whether PKA and CKI could phosphorylate Smo directly, we carried out an *in vitro* kinase assay. Three glutathione S-transferase (GST)–Smo fusion proteins containing Smo amino acids 656–755 were generated: GST–Smo^{-PKA123} has the three PKA sites mutated and GST–Smo^{-CKI} has the six CKI sites mutated to Ala (Fig. 2a). As shown in Fig. 2c, GST–Smo was phosphorylated by PKA and CKI after a primed phosphorylation by PKA (lanes 2 and 5). Mutating the PKA sites in GST–Smo abolished phosphorylation by PKA and CKI (lanes 3 and 6). Mutating the CKI sites blocked PKA-primed phosphorylation by CKI but did not affect phosphorylation by PKA (lanes 4 and 7). These results indicate that PKA phosphorylates Smo at Ser 667, Ser 687 and Ser 740, which primes its further phosphorylation by CKI.

To determine whether Hh-induced phosphorylation of Smo is mediated by PKA and CKI, S2 cells transfected with a Flag-tagged full-length Smo (Fg-Smo) were treated with Hh-conditioned

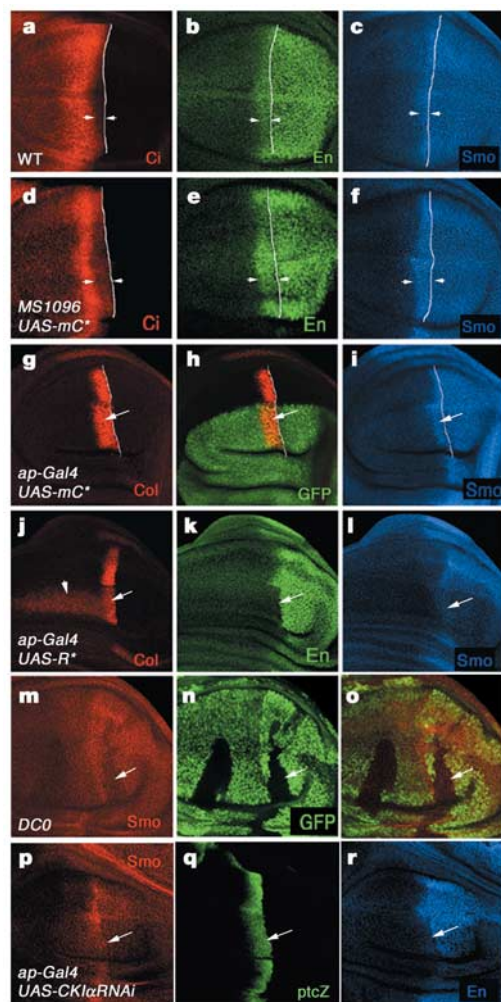


Figure 1 Positive regulation of Smo accumulation and Hh signalling by PKA and CKI. All wing discs shown in this study were oriented with anterior to the left and ventral at the top. Wild-type wing discs (**a–c**) or wing discs expressing *UAS-mC** with *MS1096* (**d–f**) were immunostained to visualize the expression of Ci155 (red), En (green) and Smo (blue). The A/P border is outlined according to the posterior limit of the Ci expression domain. **g–i**, A wing disc coexpressing *UAS-mC** and *UAS-GFP* with *ap-Gal4* and immunostained to show the expression of Col (red), GFP (green) and Smo (blue). Col and high Smo expression domains expanded anteriorly in the dorsal compartment (arrows). **j–l**, A wing disc expressing two copies of *UAS-R** with *ap-Gal4* and immunostained with anti-Col (red), anti-En (green) and anti-SmoC (blue) antibodies. Hh-induced Smo accumulation and En expression were blocked by R* (arrows in **k** and **l**). Col expression at the A/P border was attenuated (arrow in **j**). Note the low levels of ectopic Col (arrowhead in **j**), due to release of PKA inhibition on Ci. **m–o**, A wing disc carrying *DC0*⁻ clones and stained with anti-SmoC (red) and anti-GFP (green) antibodies. *DC0*⁻ clones lack GFP expression (arrows). **p–r**, A wing disc expressing *UAS-CKIαRNAi* with *ap-Gal4* was stained to show the expression of Smo (red), *ptc-lacZ* (green) and En (blue). CKIα knockdown attenuated Smo accumulation and the expression of *ptc-lacZ* and En in D-compartment cells (arrows).

medium and with or without PKA inhibitor H-89 (ref. 22) and/or CKI inhibitor CKI-7 (ref. 23). Consistent with a previous finding¹⁵, Hh induced electrophoretic mobility shift of Fg-Smo (Fig. 2d, lane 2), which is indicative of Smo phosphorylation¹⁵. Hh-induced Smo mobility shift was diminished by H-89 treatment (Fig. 2d, lane 4) and nearly abolished by a combined treatment with H-89 and CKI-7 (Fig. 2d, lane 6). Furthermore, the level of Fg-Smo

decreased when PKA or CKI activity was blocked (Fig. 2d, lanes 4–6).

To determine whether Hh induced Smo phosphorylation at PKA sites, we transfected Fg-Smo or Fg-Smo^{-PKA123} into S2 cells, followed by treatment with Hh-conditioned medium. As shown in Fig. 2e, Hh induced the appearance of slow-migrating, hyperphosphorylated forms of Fg-Smo. In contrast, Fg-Smo^{-PKA123} did

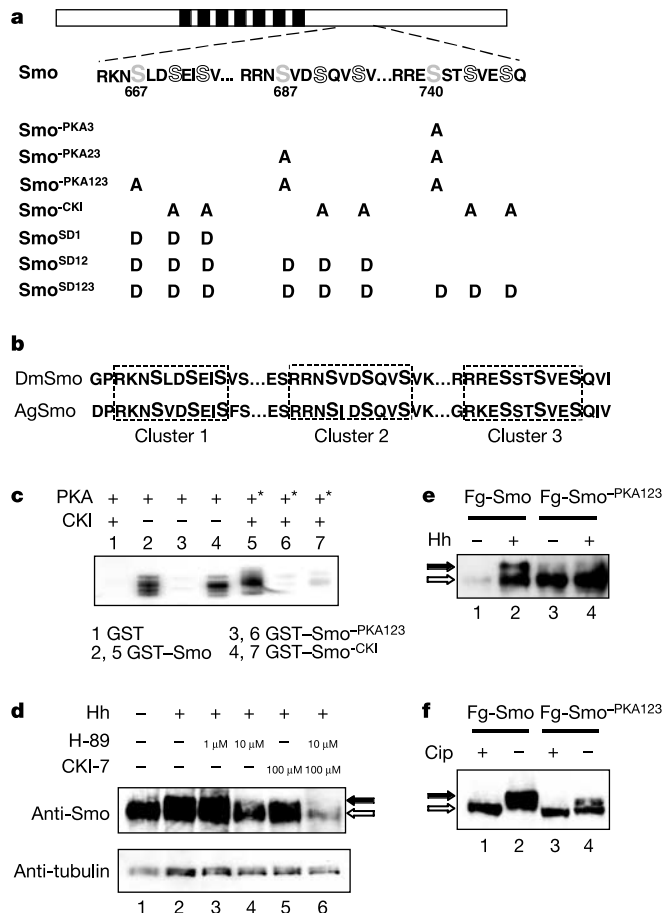


Figure 2 Phosphorylation of Smo by PKA and CKI. **a**, Full-length Smo with three phosphorylation clusters shown underneath. Large grey and open letters indicate predicted PKA and CKI sites, respectively. A list of Smo variants with corresponding substitutions is shown below. **b**, Sequence alignment of the three phosphorylation clusters between *Drosophila* Smo (DmSmo) and *Anopheles* Smo (AgSmo). **c**, Autoradiograph of an *in vitro* kinase assay. Equal amounts of GST (lane 1), GST-Smo (lanes 2 and 5), GST-Smo^{-PKA123} (lanes 3 and 6) and GST-Smo^{-CKI} (lanes 4 and 7) were incubated with recombinant PKAc in the presence of [γ -³²P]ATP (lanes 2–4), or treated with recombinant CKI in the presence of [γ -³²P]ATP (lanes 5–7) following incubation with PKA in the presence of cold ATP (indicated by an asterisk). **d**, S2 cells transfected with Fg-Smo were treated with or without Hh-conditioned medium in the presence or absence of PKA and/or CKI inhibitor at the indicated concentrations. Cell extracts were immunoprecipitated with anti-Flag antibody to enrich Smo protein, followed by immunoblotting with anti-SmoC antibody. **e**, S2 cells transfected with Fg-Smo or Fg-Smo^{-PKA123} were treated with or without Hh-conditioned medium. Cell extracts were immunoprecipitated with anti-Flag antibody and blotted with anti-SmoC antibody. **f**, Cell extracts were prepared from wing discs expressing *UAS-Fg-Smo* or *UAS-Fg-Smo^{-PKA123}* with *MS1096*, treated with or without calf intestine phosphatase (Cip), followed by immunoprecipitation with anti-Flag antibody and western blotting with anti-SmoC antibody. Cip treatment abolished mobility shift of Fg-Smo, indicating that the slow migrating forms are hyperphosphorylated. **d–f**, Solid arrows indicate hyperphosphorylated forms of Smo, open arrows indicate hypophosphorylated and unphosphorylated forms.

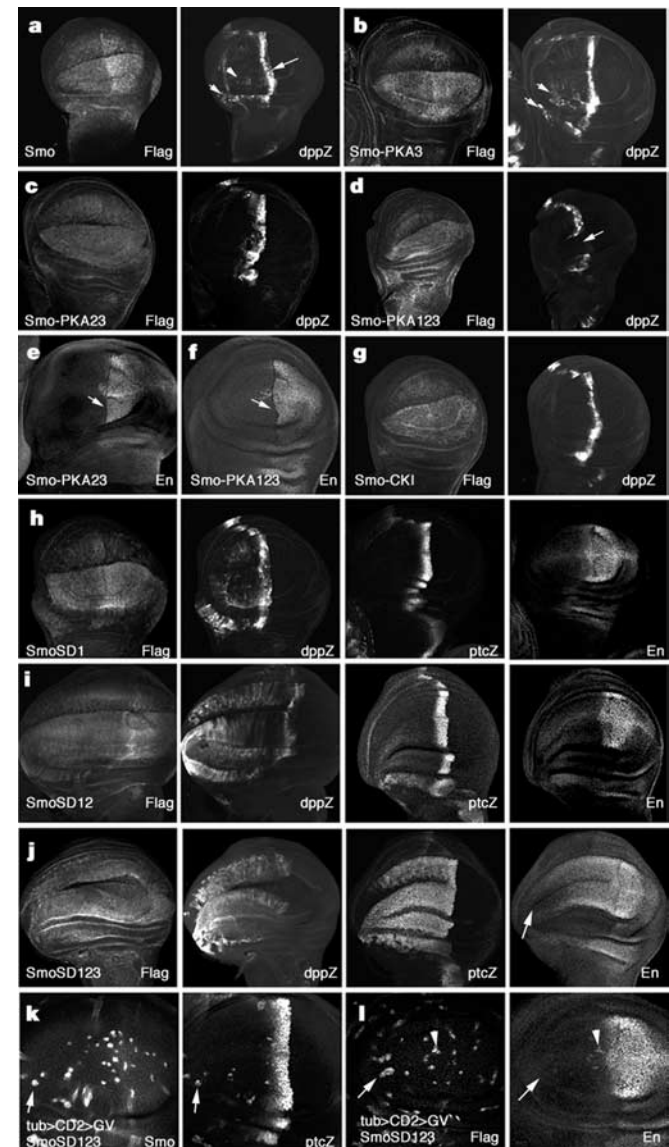


Figure 3 Hh signalling activity of Smo variants. **a–d, g**, Wing discs expressing Fg-Smo (a), Fg-Smo^{-PKA3} (b), Fg-Smo^{-PKA23} (c), Fg-Smo^{-PKA123} (d) or Fg-Smo^{-CKI} (g) with *MS1096* were immunostained to show Flag and *dpp-lacZ* expression. Arrowheads in **a** and **b** indicate the ectopic *dpp-lacZ* expression. The arrow in **a** points to the normal *dpp-lacZ* expression at the A/P border, which is blocked by Fg-Smo^{-PKA123} (indicated by the arrow in **d**). Note that the level of Gal4 driven by *MS1096* is higher in the dorsal compartment than in the ventral compartment. **e–f**, Wing discs expressing Fg-Smo^{-PKA23} (e) or Fg-Smo^{-PKA123} (f) with *MS1096* were immunostained to show En expression. Hh-induced En expression was blocked by Fg-Smo^{-PKA23} and Fg-Smo^{-PKA123} (arrows). **h–j**, Wing discs expressing Fg-Smo^{SD1} (h), Fg-Smo^{SD12} (i) or Fg-Smo^{SD123} (j) with *MS1096* were immunostained to show the expression of Flag, *dpp-lacZ*, *ptc-lacZ* and En. The arrow in **j** indicates the anterior-most region where *en* is not ectopically expressed. **k, l**, Wing discs bearing clones that express Fg-Smo^{SD123} with *tub > CD2 > GV* were immunostained to show the expression of Smo and *ptc-lacZ* (k), or Flag and En (l). Arrows (k, l) indicate clones distal to the A/P boundary whereas arrowheads (j) indicate a proximal clone.

not show significant mobility shift in response to Hh. When expressed in wing discs, Fg-Smo, which was mainly derived from P-compartment cells, existed in slow-migrating, hyperphosphorylated forms (Fig. 2f, lane 2). In contrast, Fg-Smo^{-PKA123} mostly existed in fast migrating, unphosphorylated forms (Fig. 2f, lane 4). Of note, a small fraction of Fg-Smo^{-PKA123} showed a slight mobility shift, indicating that Fg-Smo^{-PKA123} could undergo low levels of phosphorylation. Taken together, our data indicate that Hh induces phosphorylation of Smo by PKA and CKI at several sites in the Smo C-tail.

To determine whether Smo phosphorylation is essential for pathway activation, we assessed the activity of Flag-tagged wild-type or Smo variants with one, two or three PKA sites mutated to Ala (Fig. 2a). Overexpressing Fg-Smo induced ectopic, albeit weak, pathway activation, as evidenced by low levels of ectopic *dpp-lacZ* expression (Fig. 3a). Fg-Smo^{-PKA3} induced ectopic *dpp-lacZ* expression at levels comparable with Fg-Smo (Fig. 3b) and restored

Hh signalling activity in *smo*⁻ cells at the A/P border (Supplementary Fig. S1c, d). In contrast, Fg-Smo^{-PKA23} failed to induce ectopic *dpp-lacZ* expression and blocked *en* expression near the A/P border (Fig. 3c, e). However, Fg-Smo^{-PKA23} could restore *dpp-lacZ* expression and partially rescue *ptc-lacZ* expression in *smo*⁻ cells near the A/P border (Supplementary Fig. S1e, f), indicating that it could transduce low, but not high, levels of pathway activity. Fg-Smo^{-PKA123} failed to rescue *smo* mutant phenotypes (data not shown) and blocked the expression of Hh target genes, including *dpp-lacZ* and *en* (Fig. 3d, f), indicating that it not only failed to transduce Hh signal but that it also interfered with endogenous Smo. Hence, a single PKA site is necessary for low levels of pathway activity, whereas two or more PKA sites appear to be essential for high levels of pathway activity.

Fg-Smo^{-CKI} failed to induce ectopic Hh pathway activation (Fig. 3g) and did not rescue Hh target gene expression in *smo*⁻ cells (Supplementary Fig. S1g, h), indicating that phosphorylation

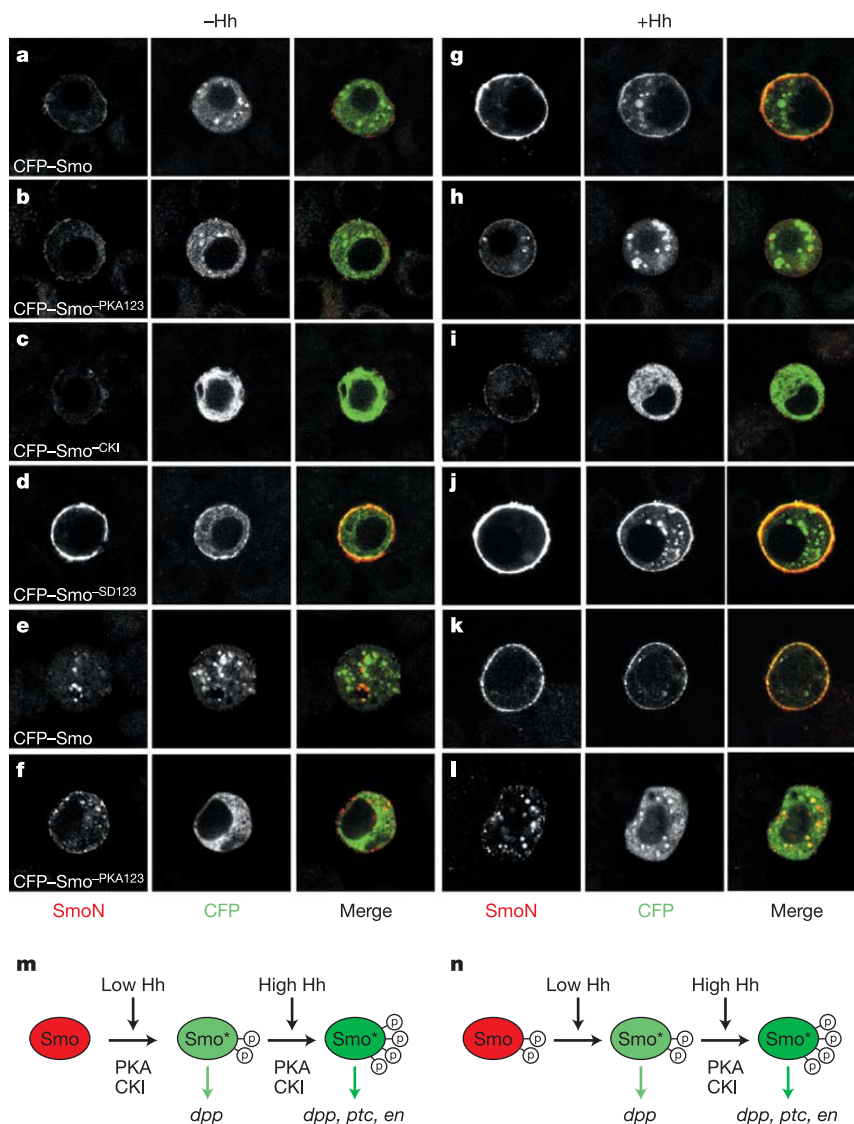


Figure 4 Regulation of Smo cell-surface accumulation by phosphorylation and models for kinase regulation of Smo activity. **a–d, g–j**, S2 cells were transfected with CFP-tagged Smo (**a, g**), Smo^{-PKA123} (**b, h**), Smo^{-CKI} (**c, i**) or Smo^{-SD123} (**d, j**), and treated with (**g–j**) or without (**a–d**) Hh-conditioned medium. Cell-surface Smo was visualized by immunostaining with anti-SmoN antibody (red channel) before cell permeabilization. **e, f, k, l**, S2 cells transfected with CFP-Smo (**e, k**) or CFP-Smo^{-PKA123} (**f, l**) and treated

with (**k, l**) or without (**e, f**) Hh-conditioned medium were incubated with anti-SmoN antibody in S2 medium to allow internalization of cell-surface-bound antibodies. Internalization of CFP-Smo but not CFP-Smo^{-PKA123} is largely inhibited by Hh. **m, n**, Models for regulating Smo activity by kinases in response to different thresholds of Hh. The asterisk indicates PKA-phosphorylation-independent activation of Smo by Hh. See text for details.

at CKI sites is also essential for Smo activity. CKI α has been implicated in regulating Ci (ref. 24). To determine whether CKI α could also regulate Smo, we blocked CKI α activity in wing discs by inherited RNA interference (RNAi) (ref. 25). Expressing a CKI α knockdown construct, *UAS-CKI α RNAi*, blocked Smo accumulation (Fig. 1p) and attenuated *ptc* and *en* expression (Fig. 1q, r). Hence, CKI α activity is required for Smo accumulation and high-threshold Hh responses.

Substitution of Ser/Thr kinase sites to acidic residues such as Asp often mimics phosphorylation. To determine whether phosphorylation of Smo by PKA and CKI suffices to promote Smo activity, we constructed several Smo variants in which PKA and CKI sites in one or more phosphorylation clusters were substituted to Asp (SD; Fig. 2a). Fg-Smo^{SD1} induced more A-compartment cells to ectopically express *dpp* at higher levels than Fg-Smo (compare Fig. 3h with a). Fg-Smo^{SD12} induced more dramatic *dpp* expression in almost all A-compartment cells and began to induce ectopic, albeit low, levels of *ptc* expression (Fig. 3i). Strikingly, Fg-Smo^{SD123} induced ectopic expression of *en* and high levels of *ptc* (Fig. 3j). The constitutive activity of Smo^{SD123} is independent of endogenous Smo, as Fg-Smo^{SD123} activated *ptc* in *smo*⁻ cells (data not shown). These observations indicate that phosphorylation of Smo at PKA and CKI sites results in constitutive Smo activity. Furthermore, the levels of Smo activity appear to correlate with its levels of phosphorylation.

Fg-Smo^{SD123} did not induce ectopic *en* expression in the anterior-most region of wing discs (Fig. 3j), implying that its activity could still be modulated by Ptc and Hh. To further address this, we expressed Fg-Smo^{SD123} using a weak Gal4 driver, *tub > CD2 > GV* (ref. 9). As shown in Fig. 3k, the level of Smo^{SD123} driven by *tub > CD2 > GV* was significantly higher than that of endogenous Smo. Under this condition, Smo^{SD123} induced modest levels of ectopic *ptc* expression (Fig. 3k) and induced ectopic *en* expression only in A-compartment cells near the A/P border where Hh is present (Fig. 3l). Hence, Ser to Asp substitutions at PKA and CKI sites confer increased activity but do not render Smo fully active, implying that Smo could be regulated by a parallel mechanism(s) independent of phosphorylation at the three clustered sites.

Fg-Smo accumulated at high levels in P-compartment cells (Fig. 3a). In contrast, Fg-Smo^{-PKA23}, Fg-Smo^{-PKA123} or Fg-Smo^{-CKI} failed to do so (Fig. 3c, d, g), indicating that Smo phosphorylation may regulate its cell-surface accumulation. To address this, we transfected S2 cells with cyan fluorescent protein (CFP)-tagged wild-type or mutant forms of Smo, with or without Hh stimulation, and stained transfected cells with anti-Smo N-terminal (SmoN) antibody before membrane permeabilization, to visualize cell-surface localized Smo. As shown in Fig. 4, CFP-Smo showed low levels of cell-surface expression in the absence of Hh (Fig. 4a), but accumulated on the cell surface upon Hh stimulation (Fig. 4g). In contrast, neither CFP-Smo^{-PKA123} nor CFP-Smo^{-CKI} accumulated on the cell surface at high levels in response to Hh (Fig. 4h, i). Strikingly, CFP-Smo^{SD123} showed high levels of cell-surface expression in quiescent cells (Fig. 4d), which may account for its constitutive activity *in vivo*. The basal levels of cell-surface CFP-Smo^{SD1} and CFP-Smo^{SD12} were also higher than those of CFP-Smo (Supplementary Fig. S2). Cell-surface expression of CFP-Smo^{SD123} was further elevated in response to Hh (Fig. 4j), indicating that Hh can regulate Smo cell-surface accumulation through a mechanism(s) in addition to phosphorylation at the three clustered sites. Similar results were obtained in wing discs (Supplementary Fig. S3).

Phosphorylation of Smo could regulate its intracellular trafficking and/or stability on the cell surface. To address this, we carried out an antibody uptake experiment. Internalized CFP-Smo was readily detected in quiescent cells (Fig. 4e), indicating that CFP-Smo could reach the cell surface but was removed by endo-

cytosis. Upon Hh stimulation, internalized CFP-Smo was greatly reduced and CFP-Smo accumulated on the cell surface (Fig. 4k). In contrast, internalized CFP-Smo^{-PKA123} was detected regardless of Hh stimulation (Fig. 4f, l), indicating that phosphorylation may regulate Smo cell-surface expression, at least in part, by inhibiting endocytosis and/or promoting recycling.

In summary, our data indicate that PKA and CKI phosphorylate Smo at several sites to promote its cell-surface accumulation and signalling activity in response to Hh. The correlation between cell-surface localization and signalling activity indicates that phosphorylation-induced cell-surface accumulation of Smo probably contributes to Hh pathway activation. Indeed, a previous study demonstrated that forced cell-surface localization of Smo stimulated Hh signalling activity, whereas endoplasmic reticulum retention of an activated form of Smo blocked its signalling activity²⁶. Interestingly, Smo^{-PKA123} not only failed to accumulate on the cell surface but also blocked the accumulation of wild-type Smo (Supplementary Fig. S4), consistent with its dominant negative effect.

Both loss- and gain-of-phosphorylation mutations show that the level of Smo signalling activity is proportional to its level of phosphorylation, indicating that graded Hh activity could be translated into graded Smo activity through differential phosphorylation. Low Hh only induces low levels of Smo phosphorylation, which suffices for low pathway activation, whereas high Hh induces high levels of Smo phosphorylation, leading to Smo cell-surface accumulation and high pathway activity (Fig. 4m). Alternatively, low Hh may not cause significant change in the levels of Smo phosphorylation, and basal phosphorylation of Smo is permissive for low pathway activity; high Hh induces increasing Smo phosphorylation, leading to amplification of pathway activity (Fig. 4n). In both models, hyperphosphorylation of Smo is essential for its cell-surface accumulation and high-threshold Hh responses.

It is unlikely that Hh induces Smo phosphorylation by increasing cyclic AMP-dependent PKA activity, as the constitutively active form of PKAc (mC*) does not induce *de novo* pathway activation. Ptc may normally block Smo phosphorylation, as knockdown of Ptc induced hyperphosphorylation of Smo (ref. 15). Ptc doesn't seem to inhibit PKA activity⁶, so it could inhibit Smo phosphorylation either by regulating a phosphatase(s), or by regulating the accessibility of Smo to its kinases. □

Methods

Mutations and transgenes

*smo*³ and *DC0*^{E95} are null alleles (refs 5, 11). *UAS-mC**, *UAS-R**, *MS1096*, *ap-Gal4*, *tub > CD2 > GV*, *hh-Gal4*, *dpp-Gal4*, *dpp-lacZ* and *ptc-lacZ* have been described^{6,9,27} (FlyBase; available at <http://flybase.bio.indiana.edu/>). Smo variants with substitutions at PKA and CKI sites were generated by site-directed mutagenesis. A Flag-tag or CFP was fused at the C terminus of each Smo variant, followed by subcloning into pUAST. To construct *UAS-CKI α RNAi*, a genomic DNA fragment with coding sequence for CKI α amino acids 153–337 was cloned into pUAST with the corresponding cDNA fragment inserted in a reverse orientation.

Generating clones of marked mutant cells

Clones of mutant cells were generated by *FLP/FRT*-mediated mitotic recombination as described⁵. Genotypes for generating clones are as follows. *DC0* clones: *y hsp-flp; DC0*^{E95} *stc FRT40/hsp-Myc-GFP FRT40*. *smo*⁻ clones expressing Smo variants: *y hsp-flp/MS1096; smo*³ *FRT 39E/hsp-CD2*, *y+ FRT39E; UAS-Fg-Smo*^{-PKA(CKI)}/*dpp-lacZ* or *ptc-lacZ*.

Cell culture, immunoblot, immunostaining and *in vitro* kinase assay

S2 cells were cultured as described²⁸. Transfection was carried out using Calcium Phosphate Transfection Kit (Specialty Media). Hh-condition medium treatment was carried out as described²⁴. We performed immunoprecipitation and western blotting analysis using standard protocols. For cell-surface staining, transfected cells were fixed with 4% paraformaldehyde and incubated with primary antibody in phosphate-buffered saline for 30 min at room temperature. For antibody uptake, transfected cells were incubated with primary antibody in S2 cell medium for 30 min at room temperature, followed by fixation, permeabilization, and secondary antibody staining. We performed immunostaining of imaginal discs with standard protocols⁵. Antibodies used in this study: rabbit anti-Ci (2A) (ref. 29), mouse anti-En (DSHB), rabbit anti-Col (ref. 4), rabbit anti-

βGal (Cappel), mouse anti-CD2 (Serotec), mouse anti-Flag (Sigma), rat anti-SmoC (ref. 15), and mouse anti-SmoN (ref. 30). For *in vitro* kinase assay, purified recombinant PKAc and CK1δ (NEB) were used.

Received 7 July; accepted 5 November 2004; doi:10.1038/nature03179.

- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Briscoe, J. & Ericson, J. The specification of neuronal identity by graded Sonic Hedgehog signalling. *Semin. Cell Dev. Biol.* **10**, 353–362 (1999).
- Strigini, M. & Cohen, S. M. A Hedgehog activity gradient contributes to AP axial patterning of the *Drosophila* wing. *Development* **124**, 4697–4705 (1997).
- Vervoort, M., Crozatier, M., Valle, D. & Vincent, A. The COE transcription factor Collier is a mediator of short-range Hedgehog-induced patterning of the *Drosophila* wing. *Curr. Biol.* **9**, 632–639 (1999).
- Jiang, J. & Struhl, G. Protein kinase A and Hedgehog signalling in *Drosophila* limb development. *Cell* **80**, 563–572 (1995).
- Li, W., Ohlmeyer, J. T., Lane, M. E. & Kalderon, D. Function of protein kinase A in hedgehog signal transduction and *Drosophila* imaginal disc development. *Cell* **80**, 553–562 (1995).
- Pan, D. & Rubin, G. M. cAMP-dependent protein kinase and *hedgehog* act antagonistically in regulating *decapentaplegic* transcription in *Drosophila* imaginal discs. *Cell* **80**, 543–552 (1995).
- Lepage, T., Cohen, S. M., Diaz-Benjumea, F. J. & Parkhurst, S. M. Signal transduction by cAMP-dependent protein kinase A in *Drosophila* limb patterning. *Nature* **373**, 711–715 (1995).
- Wang, G., Wang, B. & Jiang, J. Protein kinase A antagonizes Hedgehog signaling by regulating both the activator and repressor forms of Cubitus interruptus. *Genes Dev.* **13**, 2828–2837 (1999).
- Price, M. A. & Kalderon, D. Proteolysis of cubitus interruptus in *Drosophila* requires phosphorylation by protein kinase A. *Development* **126**, 4331–4339 (1999).
- Jia, J. *et al.* Shaggy/GSK3 antagonizes Hedgehog signalling by regulating Cubitus interruptus. *Nature* **416**, 548–552 (2002).
- Price, M. A. & Kalderon, D. Proteolysis of the Hedgehog signaling effector Cubitus interruptus requires phosphorylation by Glycogen Synthase Kinase 3 and Casein Kinase 1. *Cell* **108**, 823–835 (2002).
- Jiang, J. Degrading Ci: who is Cul-pable? *Genes Dev.* **16**, 2315–2321 (2002).
- Ohlmeyer, J. T. & Kalderon, D. Hedgehog stimulates maturation of Cubitus interruptus into a labile transcriptional activator. *Nature* **396**, 749–753 (1998).
- Defef, N., Neubuser, D., Perez, L. & Cohen, S. M. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell* **102**, 521–531 (2000).
- Ohlmeyer, J. T. & Kalderon, D. Dual pathways for induction of wingless expression by protein kinase A and Hedgehog in *Drosophila* embryos. *Genes Dev.* **11**, 2250–2258 (1997).
- Kalderon, D. & Rubin, G. M. Isolation and characterization of *Drosophila* cAMP-dependent protein kinase genes. *Genes Dev.* **2**, 1539–1556 (1988).
- Alcedo, J., Ayzenzon, M., von Ohlen, T., Noll, M. & Hooper, J. E. The *Drosophila* *smoothened* gene encodes a seven-pass membrane protein, a putative receptor for the Hedgehog signal. *Cell* **86**, 221–232 (1996).
- van-den-Heuvel, M. & Ingham, P. W. *Smoothened* encodes a receptor-like serpentine protein required for *hedgehog* signalling. *Nature* **382**, 547–551 (1996).
- Kemp, B. E. & Pearson, R. B. Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15**, 342–346 (1990).
- Umphress, J. L., Tuazon, P. T., Chen, C. J. & Traugh, J. A. Determinants on simian virus 40 large T antigen are important for recognition and phosphorylation by casein kinase I. *Eur. J. Biochem.* **203**, 239–243 (1992).
- Chijiwa, T. *et al.* Inhibition of forskolin-induced neurite outgrowth and protein phosphorylation by a newly synthesized selective inhibitor of cyclic AMP-dependent protein kinase, N-[2-(p-bromocinnamylamino)ethyl]-5-isoquinolinesulfonamide (H-89), of PC12D pheochromocytoma cells. *J. Biol. Chem.* **265**, 5267–5272 (1990).
- Chijiwa, T., Hagiwara, M. & Hidaka, H. A newly synthesized selective casein kinase I inhibitor, N-(2-aminoethyl)-5-chloroisoquinoline-8-sulfonamide, and affinity purification of casein kinase I from bovine testis. *J. Biol. Chem.* **264**, 4924–4927 (1989).
- Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
- Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nature Biotechnol.* **18**, 896–898 (2000).
- Zhu, A. J., Zheng, L., Suyama, K. & Scott, M. P. Altered localization of *Drosophila* Smoothened protein activates Hedgehog signal transduction. *Genes Dev.* **17**, 1240–1252 (2003).
- Calleja, M., Moreno, E., Pelaz, S. & Morata, G. Visualization of gene expression in living adult *Drosophila*. *Science* **274**, 252–255 (1996).
- Jia, J., Tong, C. & Jiang, J. Smoothened transduces Hedgehog signal by physically interacting with Costal2/Fused complex through its carboxyl-terminal tail. *Genes Dev.* **17**, 2709–2720 (2003).
- Motzny, C. K. & Holmgren, R. The *Drosophila* cubitus interruptus protein and its role in the *wingless* and *hedgehog* signal transduction pathways. *Mech. Dev.* **52**, 137–150 (1995).
- Lum, L. *et al.* Hedgehog signal transduction via Smoothened association with a cytoplasmic complex scaffolded by the atypical kinesin, Costal-2. *Mol. Cell* **12**, 1261–1274 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Kalderon, S. Cohen, P. Beachy, R. Holmgren and G. Struhl for reagents, and K. Wharton and H. Kramer for comments. This work was supported by grants from NIH, Leukemia and Lymphoma Society Scholar Program to J.J., who is a Eugene McDermott Endowed Scholar of Biomedical Science at UTWS.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.J. ([jin.jiang@utsouthwestern.edu](mailto:jjin.jiang@utsouthwestern.edu)).

Accurate multiplex gene synthesis from programmable DNA microchips

Jingdong Tian¹, Hui Gong¹, Nijing Sheng², Xiaochuan Zhou³, Erdogan Gulari⁴, Xiaolian Gao² & George Church¹

¹Harvard Medical School, 77 Ave Louis Pasteur, Boston, Massachusetts 02115, USA

²Department of Chemistry, University of Houston, Houston, Texas 77004-5003, USA

³Atactic Technologies Inc., 2575 West Bellfort, Suite 270, Houston, Texas 77054, USA

⁴Chemical Engineering Department, University of Michigan, Ann Arbor, Michigan 48109-2136, USA

Testing the many hypotheses from genomics and systems biology experiments demands accurate and cost-effective gene and genome synthesis. Here we describe a microchip-based technology for multiplex gene synthesis. Pools of thousands of ‘construction’ oligonucleotides and tagged complementary ‘selection’ oligonucleotides are synthesized on photo-programmable microfluidic chips¹, released, amplified and selected by hybridization to reduce synthesis errors ninefold. A one-step polymerase assembly multiplexing reaction assembles these into multiple genes. This technology enabled us to synthesize all 21 genes that encode the proteins of the *Escherichia coli* 30S ribosomal subunit, and to optimize their translation efficiency *in vitro* through alteration of codon bias. This is a significant step towards the synthesis of ribosomes *in vitro* and should have utility for synthetic biology in general.

The advance of large-scale biochemical analyses such as sequencing, microarrays and proteomics has generated vast amounts of data, which computational biologists have leveraged into a large number of hypotheses. To test these hypotheses we must first overcome the bottleneck in constructing new genetic elements, genetic pathways and engineered cells. To optimize complex biological processes using darwinian selection, we need to direct the finite diversity available in combinatorial oligonucleotide synthesis (about 25 randomized base pairs (bp) or equivalents) thoughtfully through large stretches (at the megabase level) of DNA sequence. These represent great challenges and potential payoffs for the emerging field of synthetic biology.

In principle, we can create a useful variety of molecules, cellular and cell-free systems given a sufficient supply of custom genes and genomes. However, current methods for generating even basic oligonucleotides are expensive (US\$0.11 per nucleotide) and have very high levels of errors (deletions at a rate of 1 in 100 bases and mismatches and insertions at about 1 in 400 bases). As a result, gene or genome synthesis from oligonucleotides is both expensive and prone to error. Correcting errors by clone sequencing and mutagenesis methods further increases the amount of labour and total cost (to at least US\$2 per base pair). In principle, the cost of oligonucleotide synthesis can be reduced by performing massively parallel custom syntheses on microchips^{1,2}. This can now be achieved using a variety of methods, including ink-jet printing with standard reagents (Agilent), photolabile 5′ protecting groups (Nimblegen/Affymetrix), photo-generated acid deprotection (Atactic/Xeotron) and electrolytic acid/base arrays (Oxamer/Combimatrix). However, current microchips have very low surface areas and hence only small amounts of oligonucleotides can be produced. When released into solution, the oligonucleotides are present at picomolar or lower concentrations per sequence, concentrations that are insufficiently high to drive bimolecular priming reactions efficiently.

A potential solution to this scale problem would be to amplify the

oligonucleotides obtained from the microchips from roughly as little as 10^5 (or 10^9 for low density arrays) up to 10^9 (or 10^{12}) molecules of each sequence, thereby permitting subsequent selection and assembly steps. An overview of the integrated process is presented in Fig. 1.

For this amplification method (shown in Fig. 2a), oligonucleotides flanked by short, generic adaptor sequences are synthesized on a programmable microchip. This generates a pool of 10^2 – 10^5 different oligonucleotides, which can be released from the microchips by chemical or enzymatic treatment. Released oligonucleotides are amplified by polymerase chain reaction (PCR) using primers that contain type-IIS restriction enzyme recognition sites. Digestion of the PCR products with the corresponding restriction enzyme(s) yields sufficient amounts of unadulterated oligonucleotide sequences to be used for gene or genome assembly.

We first demonstrated the feasibility of this approach with Atactic/Xeotron 4K (that is, 3,968 synthesis chambers) photo-programmable microfluidic microarrays¹. To monitor oligonucleotide synthesis and cleavage from the microchip, the 5' ends of the oligonucleotides were coupled with fluorescein. The microchip was scanned with a microarray scanner before and after cleavage, and the images are shown in Fig. 2b. The cleaved portions of the oligonucleotides were hybridized onto a 'quality-assessment (QA)-chip' synthesized with complementary oligonucleotide sequences (Fig. 2c). These results demonstrated that individual oligonucleotides were synthesized and nearly completely released from the microchip in quantities that can be measured by a QA-chip hybridization process. The typical yield of oligonucleotide released from each chamber of the 4K microchip is about 5 fmoles, as determined by quantitative PCR¹. Using primers that anneal specifically to the generic adaptors flanking the oligonucleotide sequences, PCR reactions were carried out to amplify the oligonucleotides more than a million-fold.

Mutations incurred during oligonucleotide synthesis are a major source of errors in assembled DNA molecules, and are costly and difficult to eradicate^{4,5}. We developed a simple, stringent hybridization-based method to remove oligonucleotides with

such mutations. To select against mutations in gene construction oligonucleotides, these oligonucleotides were hybridized sequentially to two pools of bead-immobilized short complementary selection oligonucleotides that together span the entire length of the construction oligonucleotides (Fig. 3a). All selection oligonucleotides were designed to have nearly identical melting temperatures by varying their lengths. Under appropriate hybridization conditions, imperfect pairs between selection and construction oligonucleotides due to base-mismatch or deletion have lower melting temperatures and are unstable. After the cycles of hybridization, wash and elution, oligonucleotides with sequences that perfectly match the selection oligonucleotides are preferentially retained and enriched. Digestion of the PCR products with type-IIS restriction enzymes removed the generic primer sequences from both ends of the oligonucleotides (Fig. 3b, lanes 1–2). In these experiments the amplification tags are removed just before selection. However, if the digestion were deferred, the oligonucleotides could be re-amplified by PCR and subjected to further rounds of hybridization selection. Because the probability of complementary mutations occurring at matching positions on construction and selection oligonucleotides is miniscule, in principle most oligonucleotides with mutations can be eliminated by this selection procedure.

Like construction oligonucleotides, selection oligonucleotides were also synthesized and released from programmable microarrays. Selection oligonucleotides with arms were amplified by PCR, and the strands complementary to the gene construction oligonucleotide were labelled with biotin at the 5' end and selectively immobilized on streptavidin beads. The unlabelled strands were denatured and removed. As demonstrated in Fig. 3b (lane 3), immobilized selection oligonucleotides selectively retained the correct 50-bp construction oligonucleotides.

Such oligonucleotides are suitable for gene assembly. To facilitate automation, we developed a single-step polymerase assembly multiplexing (PAM) reaction for multiple gene syntheses from a single pool of oligonucleotides. Single-fragment assembly methods

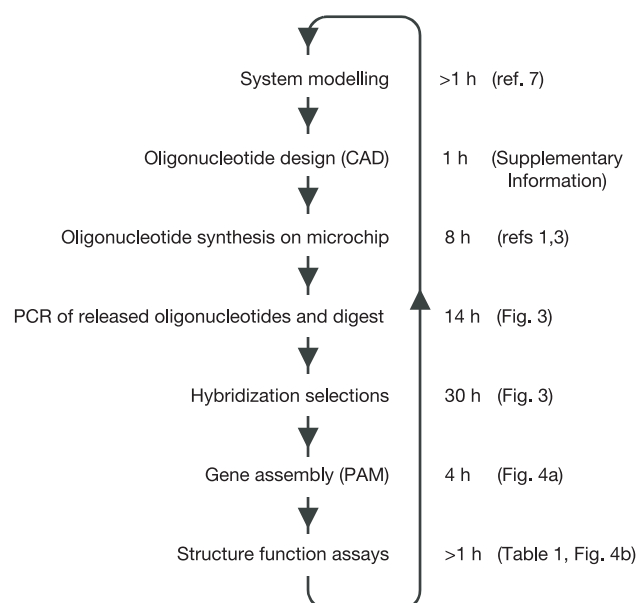


Figure 1 A flow chart for the design, synthesis and analysis of multiple genes in pools. Estimates of the current process timing (not always the minimum possible times) and sources of additional details are listed.

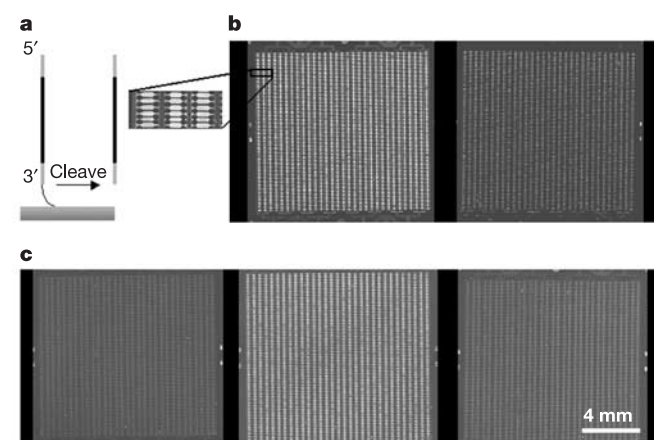


Figure 2 Preparation of free oligonucleotides from a custom microarray. **a**, Diagram of synthesis and cleavage of a PCR-amplifiable oligonucleotide from a microchip surface. The portion of the oligonucleotide used for gene construction is in black; PCR primer-adaptors are shown in grey. **b**, Synthesis and cleavage of oligonucleotides from a Xeotron/Atactic 4K photo-programmable microfluidic microchip. Left: fluorescent scanning micrograph of an oligonucleotide array before cleavage. Insert: details of microfluidic chambers and connecting channels. Right: array after cleavage. **c**, Hybridization of released fluorescein (FAM)-labelled oligonucleotides to a QA-chip. Left: before hybridization; middle: after hybridization; right: after stripping of hybridized oligonucleotides.

Table 1 Comparison of sequence errors generated by various methods

Method	Total bp	Transition	Transversion	Deletion	Addition	bp per error
Hybridization selection (PAM)	23,641	7	3	5	2	1,394
PAGE selection (PAM)	24,546	28	12	11	3	455
No selection (PAM)	9,243	25	13	19	1	159
No selection (ligation)	6,093	6	6	22	4	160

We performed χ^2 tests for hybridization selection versus PAGE selection ($P = 2 \times 10^{-5}$), and hybridization selection versus no selection ($P = 2 \times 10^{-21}$). Only the constructs in the row labelled 'PAGE selection' involved gel purification.

have traditionally used two or three steps (ligation, assembly and PCR)^{4–6}. For PAM, gene-flanking primer pairs were added to the pool of gene-construction oligonucleotides (with the primer pairs at a higher concentration than the oligonucleotides), together with thermostable polymerase and dNTPs. Extension of overlapping oligonucleotides and subsequent amplification of multiple full-length genes were accomplished in a closed-tube, one-step reaction using a thermal cycler. Different generic adaptor sequences can be incorporated into the ends of each gene or gene set, and a set of complementary adaptor-primer pairs can be pre-synthesized to avoid the cost of synthesizing gene-specific PAM primer pairs and to facilitate automation (for example, 96 or 384 generic adaptors to match standard multi-well plates).

To determine the efficiency of the hybridization-selection method to eliminate mismatch mutations⁷, we constructed genes using the same pool of microchip-synthesized oligonucleotides purified in three different ways: unpurified, polyacrylamide gel electrophoresis (PAGE)-purified or hybridization-purified. These genes were cloned and random clones from each category were sequenced in both directions to determine error types and rates for each category. As shown in Table 1, genes synthesized with unpurified oligonucleotides have the highest error rates (1 in 160 bp); the method of gene assembly (using ligation or PAM) made little difference. PAGE purification of oligonucleotides reduced the error rate to 1 in 450 bp, mainly through removal of deletion mutations. This rate is comparable to figures reported by other groups using PAGE purification^{4,5}. With hybridization selection, the error rate was further reduced to approximately 1 in 1,394 bp.

As an example of the usefulness of this technology for large-scale synthetic biology projects⁸, we used a microchip to redesign and

synthesize codon-altered versions of the 21 protein-encoding genes that constitute the *E. coli* small ribosomal subunit. We had previously noticed while using the natural versions of these 21 proteins⁹ that translational efficiencies were very low *in vitro*, even though *in vivo* the proteins have high expression levels. Redesigning codon usage is a way to increase protein translation efficiencies, although it is more challenging to accomplish when starting with nearly ideal codons. Because many other proteins are expressed well in this *in vitro* system, we hypothesized that some of the problem was due to secondary structure (possibly exacerbated by the fact that the rate of T7 polymerase-mediated transcription is eightfold higher than translation^{10,11}). We attempted to replace codons with sequences likely to have less secondary structure (for example, by lowering G + C content). Our computer-aided design software (CAD-PAM) designed overlapping 50-bp oligonucleotide sequences (embedded in 70-mers) for the 21 ribosomal genes and synthesized them all on a 4K Xeochip. These oligonucleotides were processed and hybridization-selected with selection oligonucleotides, and were then used to construct the 21 ribosomal genes in multiple PAM reactions (Fig. 4a). Error-free clones were tested in *E. coli* using coupled *in vitro* transcription–translation reactions. The translation profiles of the synthetic genes are shown in Fig. 4b. A number of codon-altered genes had higher translation levels in the *E. coli* extract compared with their respective wild-type genes. We next combined these 21 genes using sequential PAM reactions to give a pool of ~14.6 kb assemblies (Fig. 4c) by introducing unique ~30-mer overlapping linkers between gene units and performing sequential PAM reactions. Correct assembly was confirmed by sequencing on average four individual clones from every over-

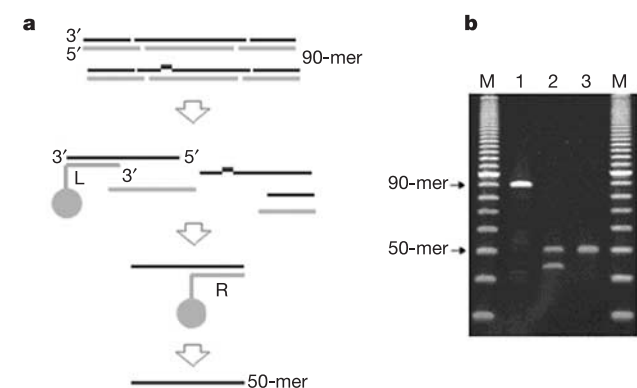


Figure 3 Hybridization selection of microchip-synthesized oligonucleotides. **a**, Diagram of the hybridization selection strategy. 90-mer oligonucleotides (upper strands black, lower strands grey) are cut with type-IIS restriction enzymes to release hybrids of 50-mers and complementary 44-mers, some of which have incorrect sequences (indicated by a bulge in the upper strand of the second 90-mer oligonucleotide). Only the correct upper 50-mer strand hybridizes well with left (L) then right (R) selection oligonucleotides (immobilized on beads in grey). **b**, A denaturing PAGE gel separating PCR-amplified oligonucleotides (90-mer, lane 1), *BsaI/BseRI*-digested oligonucleotides (50- and 44-mer, lane 2) and hybridization-selected oligonucleotides (50-mer, lane 3). Ten-base-pair markers are shown in lanes marked M.

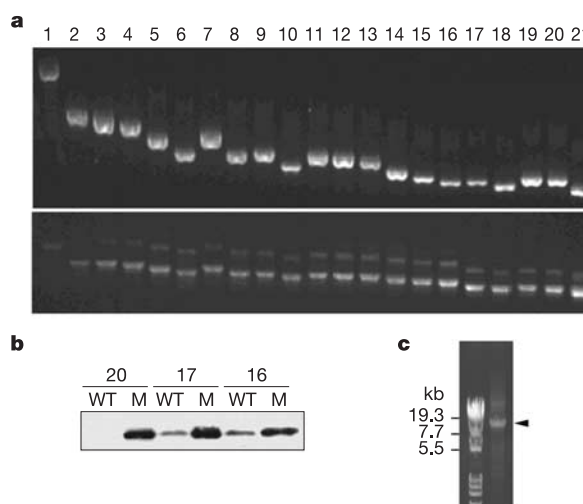


Figure 4 Synthetic gene constructs and their translation products. **a**, Twenty-one engineered ribosomal protein genes constructed using PAM reactions (upper gel) and their *in vitro* transcribed RNAs (lower set of bands in the lower gel). **b**, Ribosomal proteins translated from their synthetic genes. Western blot comparison of translation efficiency between wild-type (WT) and engineered (M) versions of three ribosomal protein genes. **c**, A 14.6-kb operon of 21 ribosomal genes assembled using sequential PAM reactions. Left lane, DNA molecular weight marker IV (Roche).

lapping DNA segment generated by high-fidelity PCR reactions that together covered the whole construct. By starting with correct input gene sequences, and through repeated high-fidelity, polymerase-based extension reactions, the assembly process resulted in a lower error rate (about 1 in 7,300 bp) than any of the methods shown in Table 1 (all of which started with oligonucleotides containing synthetic errors). This clearly demonstrated that a major source of error for gene assembly comes from oligonucleotide chemical synthesis rather than polymerase proofreading activity. Although the increasing length of the PCR products might be expected to reduce yield in the later assemblies, the number of reaction components decreases and so the efficiency remains high. Eventually when PAM length does become limiting, homologous recombination might allow assembly in the megabase range.

In this demonstration, a small fraction of a microchip's synthetic capacity was used. Massively parallel microchip-based oligonucleotide synthesis might increase yields in oligonucleotide synthesis from 9 bp per US\$ to 20 kb per US\$, depending on the type of microchip used and the number of oligonucleotides on a microchip. To achieve few or no errors in megabase-scale assemblies, the progress here on mismatch errors must be extended. In addition to the physical methods described here, we and others are using protein-based methods (for example, using MutS¹² or MutHLS mismatch correction proteins) already shown to be capable of exceeding by tenfold the fidelity of DNA polymerases (the latter being 6×10^{-6} errors per base per cycle)¹³. This could significantly reduce the money spent on sequencing and potentially eliminate the need for error correction using directed-mutagenesis methods. Ultra-low-cost DNA sequencing¹⁴ could also have a large impact if properly integrated. The next stage in testing the limits to simultaneous synthesis and assembly will employ 95,000–382,000 oligonucleotides per \$700 microchip from Nimblegen (yielding 2–18 Mbp). The first such assembly has been successful (see <http://arep.med.harvard.edu/SBP/>). Overcoming the challenge of repeated sequences in desired synthetic constructs may require approaches like hierarchical assembly, assembly focused on only the unique ends, and/or multiplex size selections of assembly products. Using geometrical constraints to force a certain order of assembly illustrates another potential strength of microfluidic syntheses. Improvement of error rates enables a variety of applications without cloning or sequencing.

Although oligonucleotide assembly is an established procedure^{6,15,16}, our procedures for elution from DNA microchips, oligonucleotide amplification, mismatch error correction, multiplexing and microfluidic integration are new. Accurate, low-cost multiplex gene syntheses will be a powerful tool for synthetic biology⁸ and complex nanostructures in general¹⁷. For example, this technology is currently enabling us to improve and test components needed for the synthesis of ribosomes *in vitro*. The rapid prototyping of individual genes on short linear templates has greatly aided the debugging of this complex system. Our ability to remap a whole set of genes from the *Mycoplasma*¹⁸ codons to those of *E. coli* (for example, by eliminating the UGA 'stop' codon and changing G + C content from 25% to 51%) will help us to calibrate proteomics experiments, test *de novo* protein designs and identify new biochemical activities, including those that are missing from our list of the core components of novel self-replicating systems⁸. □

Methods

Design of sequences

Gene and oligonucleotide sequences were designed using the Java program CAD-PAM, to be described in detail elsewhere (J.T., H.G. and G.C., manuscript in preparation). Basically, CAD-PAM uses constraints on the amino acid sequences, codon usage, messenger RNA secondary structure and restriction enzymes used to release the construction oligonucleotides in order to create nearly optimal, overlapping sets of *n*-mer (typically 50-mer) construction oligomers and shorter selection oligomers (typically 26-mer). The melting temperatures (T_m) of overlapping regions between adjacent gene construction

oligonucleotides or between construction and selection oligonucleotides were equalized. The selection oligonucleotides were padded with extra adenine residues to keep oligomer length constant (70-mers) for optional size selection (not used for typical PAM). T_m values were calculated using the nearest neighbour method¹⁹. Codons can be fixed or altered to allow expression improvements. The sequences of the oligonucleotides used in this study are provided in Supplementary Table 1.

Microchip synthesis, amplification and selection of oligonucleotides

Oligonucleotides were synthesized on photo-programmable microfluidic microchips with a phosphate at the 5' end and the 3' end coupling to the 3'-hydroxy terminus of a uracil residue (X. Gao, unpublished work). After synthesis, the oligonucleotides were cleaved either with RNase A or by ammonium hydroxide treatment (used for deprotection as in standard oligonucleotide syntheses) followed by precipitation. Gene construction oligonucleotides that had been PCR amplified with 20-mers (initially complementary to the terminal ten bases) were digested with the type-IIS restriction enzymes *BsaI* and *BseRI* (without gel purification except for the 'PAGE' controls in Table 1). While this paper was in review a related method using the enzyme *MlyI* appeared²⁰. Immobilization of biotin-labelled selection oligonucleotides on magnetic streptavidin beads (Dynal) and removal of the non-biotinylated strand were done as described²¹. Construction oligonucleotides were denatured at 95 °C for 3 min and hybridized to selection oligonucleotides in hybridization buffer (5 × SSPET buffer, 50% formamide, 0.2 mg ml⁻¹ BSA) for 14–16 h at 42 °C on a rotor. Beads were washed three times with 0.5 × SSPET and three times with wash buffer (20 mM Tris-HCl pH 7.0, 5 mM EDTA, 4 mM NaCl) at room temperature. The construction oligonucleotides were recovered by denaturation in 0.1 M NaOH for 15 min and subsequent neutralization.

Polymerase assembly multiplexing reactions

PAM reactions were carried out in 25-μl reactions containing 2 μl of oligonucleotide mixtures, 0.4 μM of each of the gene-end primer pairs, 1 × dNTP mixture and 0.5 μl of Advantage 2 polymerase mixture in 1 × buffer (Clontech Advantage 2 PCR kit). Samples were denatured at 95 °C for 3 min, then underwent 40–45 thermal cycles of 95 °C for 30 s, 49 °C for 1 min and 68 °C for 1 min kb⁻¹, then finished at 68 °C for 10 min. Sequential PAM reactions were used to combine multiple genes. First, His6-tagged linear expression constructs of the correct sequences of 21 ribosomal protein genes were pre-constructed by PCR using an RTS *E. coli* linear template generation kit (Roche). These constructs were then used as templates in separate PCR reactions where unique ~30-mer linkers with identical T_m (0.4 μM of each, Integrated DNA Technologies, Inc.) were introduced to create enough overlapping sequences between genes for secondary PAM reactions (see Supplementary Table 2). In these, three large fragments were made in separate Roche Expand long template PCR reactions: RS1–5 (1–5,513), RS6–13 (5,483–10,526) and RS14–21 (10,497–14,593). These fragments were gel-purified and assembled into a full 14,593-bp operon in the final assembly reaction using RS1–21 (1–14,593). For the last two assemblies, samples were denatured at 92 °C for 2 min, followed by 10 thermal cycles at 92 °C for 30 s, 65 °C for 1 min and 68 °C for 1 min kb⁻¹, then followed by 25 additional cycles at 92 °C for 30 s, 65 °C for 1 min, and 68 °C for 1 min kb⁻¹ plus 10 s per cycle, and finished at 68 °C for 10 min.

Coupled *in vitro* transcription and translation

Assembled genes were cloned and error-free clones were selected by sequencing. Linear constructs for *in vitro* protein expression were made using Roche RTS *E. coli* linear template generation set, His-tag. *In-vitro*-coupled transcription and translation was performed using a Roche Rapid Translation System RTS 100 *E. coli* HY kit. Proteins were detected by western blotting with an anti-His6-peroxidase antibody (Roche) using standard procedures.

Received 26 April; accepted 22 October 2004; doi:10.1038/nature03151.

1. Zhou, X. *et al.* Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.* **32**, 5409–5417 (2004).
2. Fodor, S. P. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
3. Gao, X. *et al.* Flexible DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res.* **29**, 4744–4750 (2001).
4. Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
5. Smith, H. O., Hutchison, C. A. III, Pfannkuch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA* **100**, 15440–15445 (2003).
6. Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides. *Gene* **164**, 49–53 (1995).
7. Eason, R. G. *et al.* Characterization of synthetic DNA bar codes in *Saccharomyces cerevisiae* gene-deletion strains. *Proc. Natl Acad. Sci. USA* **101**, 11046–11051 (2004).
8. Forster, A. C. & Church, G. M. A synthetic biology project. *Nature* (submitted).
9. Culver, G. M. & Noller, H. F. Efficient reconstitution of functional *Escherichia coli* 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5**, 832–843 (1999).
10. Iost, L., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes *in vivo*. *J. Bacteriol.* **174**, 619–622 (1992).
11. Iost, L. & Dreyfus, M. mRNAs can be stabilized by DEAD-box proteins. *Nature* **372**, 193–196 (1994).
12. Carr, P. A., Park, J. S., Lee, Y.-J., Yu, T., Zhang, S. & Jacobson, J. M. Protein-mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res.* (in the press).
13. Smith, J. & Modrich, P. Removal of polymerase-produced mutant sequences from PCR products. *Proc. Natl Acad. Sci. USA* **94**, 6847–6850 (1997).

14. Shendure, J., Mitra, R., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nature Rev. Genet.* **5**, 335–344 (2004).
15. Mullis, K. *et al.* Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **LI**, 263–273 (1986).
16. Dillon, P. J. & Rosen, C. A. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *BioTechniques* **9**, 298–300 (1990).
17. Liu, D., Park, S. H., Reif, J. H. & LaBean, T. H. DNA nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. *Proc. Natl Acad. Sci. USA* **101**, 717–722 (2004).
18. Jaffe, J. D., Berg, H. C. & Church, G. M. Proteogenomic mapping reveals genomic structure and novel proteins undetected by computational algorithms. *Proteomics* **4**, 59–77 (2004).
19. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA* **83**, 3746–3750 (1986).
20. Richmond, K. E. *et al.* Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32**, 5011–5018 (2004).
21. Espelund, M., Stacy, R. A. & Jakobsen, K. S. A simple method for generating single-stranded DNA probes labeled to high activities. *Nucleic Acids Res.* **18**, 6157–6158 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements DARPA BioComp and DOE GTL provided support for J.T., H.G. and G.C. We thank G. Culver, T. Wu, T. Forster, P. Carr, J. Jacobson and other members of the synthetic biology community for advice; N. Novikov for technical assistance, and E. Nuwaysir and T. Albert for help in designing the Nimblegen arrays. J.T. was supported by a LSRF fellowship. X.Z., E.G. and X.G. thank the NIH, DARPA and the R.A. Welch Foundation for grants.

Competing interests statement The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to G.C. (g1m1c1@arep.med.harvard.edu). The 14.6-kb operon of 21 ribosomal genes has been deposited in GenBank under accession number AY773199.

The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes

Joseph T. Wade, Daniel B. Hall & Kevin Struhl

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Ribosomal protein (RP) genes in eukaryotes are coordinately regulated in response to growth stimuli and environmental stress, thereby permitting cells to adjust ribosome number and overall protein synthetic capacity to physiological conditions^{1–5}. Approximately 50% of RNA polymerase II transcription is devoted to RP genes⁵. The transcriptional regulator Rap1 binds most yeast RP promoters⁶, and Rap1 sites are important for coordinate regulation of RP genes^{7–10}. However, Rap1 is not the specific regulator that controls RP transcription because it also functions as a repressor, and many Rap1-activated promoters are not coordinately regulated with RP promoters^{11,12}. Here we show that the transcription factors Fhl1 and Ifh1 associate almost exclusively with RP promoters; association depends on Rap1 and (to a lesser extent) a DNA element at many RP promoters. Ifh1 is recruited to promoters via the forkhead-associated (FHA) domain of Fhl1; the level of Ifh1 associated with RP promoters determines the level of transcription; and environmental stress causes a marked reduction in the association of Ifh1, but not Fhl1 or Rap1. Thus, Ifh1 association with promoters is the key regulatory step for coordinate expression of RP genes.

Yeast cells contain ~150 copies of the ribosomal DNA locus and 137 RP genes that encode one or two copies of each of 78 proteins. Transcription of the RP genes is coordinately regulated in accord with the cellular growth rate in a manner that requires protein kinase A and the TOR pathway^{8,13–15}. In addition, RP transcription is rapidly and coordinately downregulated in response to a variety of

environmental insults such as heat shock, amino-acid starvation and osmotic shock. Coordinate regulation of RP genes has a major biological impact on the overall protein synthetic capacity and growth of the cell, and 50% of RNA polymerase II transcription in yeast is devoted to RP genes⁵.

Despite the biological importance of coordinate regulation of RP genes, information about the mechanism of this regulation is rather incomplete. Approximately 90% of RP promoters contain predicted Rap1 binding sites^{5,16}, and Rap1 is bound to essentially all such RP promoters *in vivo*⁶. Rap1 sites are important for growth-regulated expression of natural RP promoters^{7–9}, and a 41-base-pair (bp) fragment containing two Rap1 sites from an RP promoter is sufficient to mediate growth regulation¹⁰. However, Rap1 also binds and activates many non-RP promoters that are not regulated in the same manner as RP promoters. Furthermore, the Rap1-dependent activator that regulates transcription of RP genes recruits TFIID, whereas the Rap1-dependent activator that regulates transcription of glycolytic genes does not¹⁰. Taken together, these results strongly suggest the existence of an unknown protein(s) that is specifically involved in the coordinate regulation of RP genes.

An extensive genome-wide analysis of target sites for over 100 yeast DNA-binding transcription factors revealed Fhl1, a protein containing a fork head DNA-binding domain¹⁷, as binding specifically to RP promoters¹⁸. We confirmed this result by performing chromatin immunoprecipitation (ChIP) on Fhl1 coupled with analysis on microarrays containing essentially all yeast promoter regions (Fig. 1a; Supplementary Table 1). Using a stringent cutoff of fivefold enrichment, we identified 79 targets, 76 of which are RP promoters. At this level of stringency, Fhl1 is almost exclusively specific to RP promoters. It is highly likely that Fhl1 interacts with additional RP promoters, as 80% (103 out of 129) of detectable RP promoter regions are found in the top 5% of the Fhl1-bound targets (Fig. 1a).

We also determined the genome-wide association of Ifh1, a protein that interacts genetically with Fhl1 (ref. 19). The binding profile of Ifh1 is strikingly similar to that of Fhl1, located almost exclusively at RP promoters (Fig. 1a; Supplementary Table 1). Of the 56 targets showing at least fivefold enrichment, 54 were RP promoters, and it is likely that Ifh1 associates with additional RP promoters (Fig. 1a). Remarkably, 94% (51 out of 54) of Ifh1 targets defined in this manner are also Fhl1 targets. Fhl1 and Ifh1 do not bind to non-RP, Rap1-containing promoters or to the ribosomal DNA loci (Fig. 1a, b; data not shown). Ifh1 and Fhl1 do not associate with eight of the nine RP promoters that do not bind Rap1 *in vivo*^{6,16}. Thus, Fhl1 and Ifh1 bind almost exclusively to a common set, but not all RP promoters, and binding of these proteins seems to be influenced by Rap1.

Conserved DNA sequence motifs among Fhl1- and Ifh1-bound promoters include the Rap1 binding site, an A-rich stretch implicated in RP transcription²⁰, and a close match to a sequence element, 'Motif 213', recently identified in a subset of RP promoters by a computational analysis designed to predict gene expression using DNA sequence information²¹ (Fig. 2a; Supplementary Fig. 1). Analysis of RP promoters that do not display fivefold enrichment of Fhl1 and Ifh1 identifies the Rap1 site and the A-rich stretch, but not the third motif. We therefore name the third motif IFHL, and a search of this motif across the yeast genome shows that it is highly over-represented at RP gene promoters ($P < 0.001$). Although the presence of the IFHL motif correlates with binding of Ifh1 and Fhl1 *in vivo*, there is not a strict one-to-one correspondence because 38 of the Fhl1 targets do not contain a clear IFHL motif.

Mapping of Fhl1 and Rap1 binding sites across three RP promoters indicate promoter-specific differences in the relative locations of these proteins (Fig. 2b–d). At *RPL12A*, Rap1 and Fhl1 associate with discrete regions, with Rap1 binding near the two predicted Rap1 sites and Fhl1 binding near the IFHL motif (Fig. 2b). In contrast, Fhl1 binding at *RPS11B* and *RPL40A* matches that of

Rap1 at the upstream boundary, but it seems to bind to a broader region that spans the Rap1 sites and the IFHL motif (Fig. 2c, d). This unusual profile of Fhl1 binding at these promoters is significant because in the identical samples Rap1 is localized near the predicted Rap1 sites.

Rap1, Fhl1 and Ifh1 bind a promoter derivative comprising the *RPL20B* upstream region (contains a single Rap1 site and a single IFHL motif), the *PGK1* core promoter and the *HIS3* coding region (Fig. 2e). Mutation of the Rap1 site markedly reduces binding of Rap1, Fhl1 and Ifh1, and transcriptional activity. Mutation of the IFHL motif results in a smaller but significant decrease in the

binding of Fhl1 and Ifh1, and it also causes a significant decrease in transcription. These observations indicate that Rap1 is required for Fhl1 and Ifh1 binding and that the IFHL motif contributes to binding of Fhl1 and Ifh1. In addition, they suggest that Fhl1 and Ifh1 are required for maximal transcription from an RP promoter.

The mutational analysis and the promoter-specific Fhl1 binding profiles are interesting in connection with the observation that Fhl1 and Ifh1 bind poorly at RP promoters lacking Rap1 sites, but can bind well at a number of RP promoters lacking a good match to the IFHL motif. RP promoters often contain multiple Rap1 sites or a single Rap1 site together with an element strongly resembling an

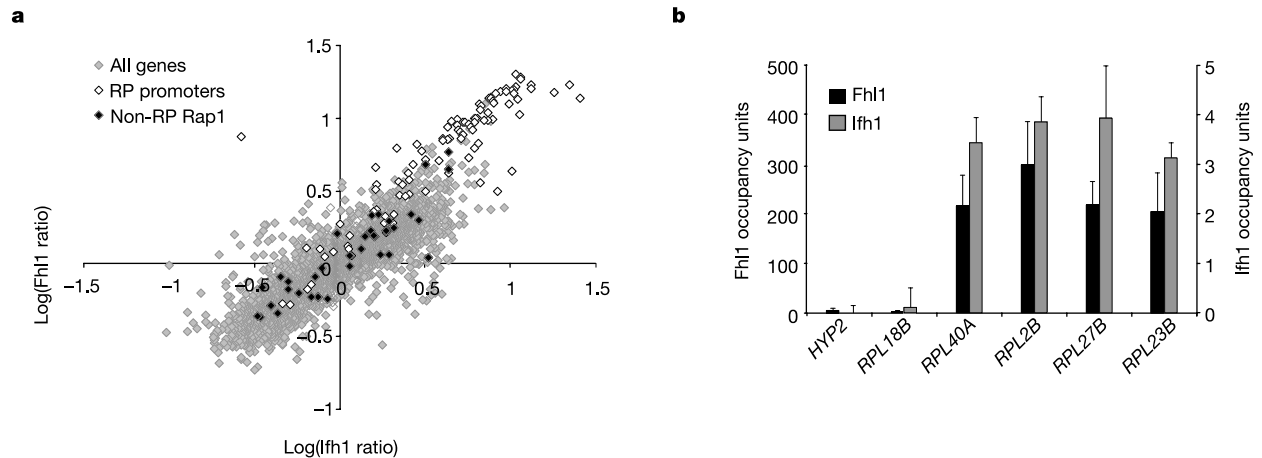


Figure 1 Promoter-specificity of Fhl1 and Ifh1. **a**, Comparison of Fhl1 and Ifh1 ChIP-chip analyses. Log ratios for Fhl1 plotted against log ratios for Ifh1 for every gene represented on the microarrays. Data for RP promoters are shown as white diamonds. Data for Rap1-bound non-RP promoters are shown as black diamonds. **b**, Fhl1 and Ifh1

occupancies at the indicated promoters. Note that *RPL18B* does not show binding of Fhl1 or Ifh1 by microarray analysis and is not bound by Rap1. Error bars reflect the standard deviation of the mean.

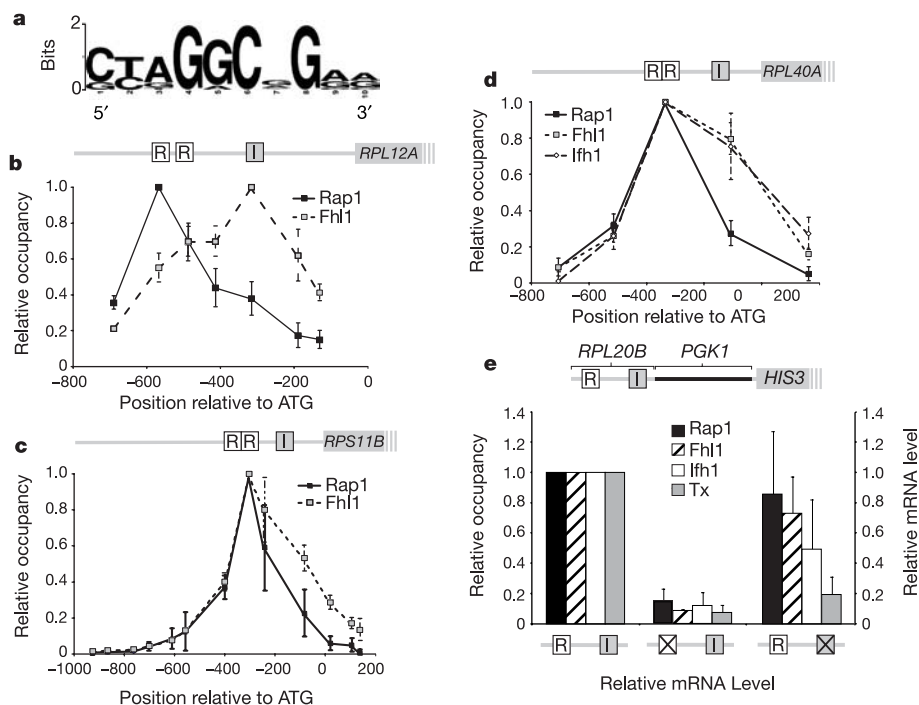


Figure 2 Role of the Rap1 site and the IFHL motif. **a**, The IFHL motif as derived from the Fhl1 ChIP-chip analysis. **b–d**, Occupancy of Rap1 and Fhl1 at various positions across the *RPL12A* (**b**), *RPS11B* (**c**) and *RPL40A* (**d**) promoters, normalized to the highest occupancy value. The positions of predicted Rap1 DNA sites and IFHL motifs are indicated above each graph by 'R' and 'I', respectively. **e**, Relative *HIS3* mRNA levels (Tx; normalized to the wild-

type promoter) and Rap1, Fhl1 and Ifh1 occupancy at *RPL20B* UAS-*PGK1* core promoter-*HIS3* open reading frame (ORF) derivatives. Wild-type *RPL20B* UAS (R-I) and derivatives containing mutations in either the Rap1 site (X-I) or the IFHL motif (R-X). Error bars reflect the standard deviation of the mean.

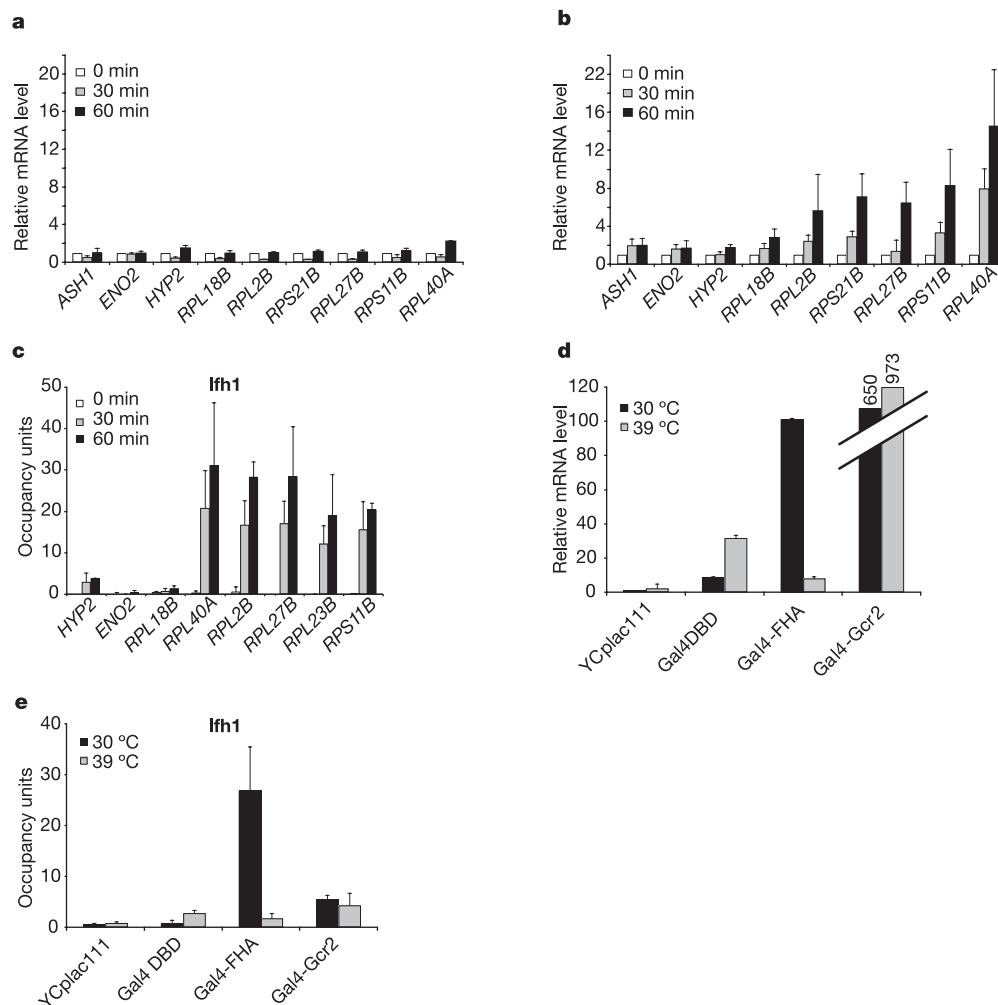


Figure 3 Ifh1-dependent induction of transcription at Ifh1-bound RP genes. **a**, Relative RNA levels of indicated genes 0, 30 and 60 min after galactose addition in a strain expressing a chromosomal copy of *IFH1* under the control of the *IFH1* promoter. **b**, Relative RNA levels of indicated genes 0, 30 and 60 min after galactose addition in a strain expressing a plasmid-borne copy of *IFH1* under the control of the *GAL1* promoter.

c, Occupancy of Ifh1 at the indicated promoters in the same cells. **d**, *GAL7* mRNA levels in cells containing either empty vector (YCplac111), Gal4 DBD, Gal4-FHA or Gal4-Gcr2, before and after a heat shock. **e**, Ifh1 occupancy at the *GAL7* promoter in the same cells. Error bars reflect the standard deviation of the mean.

IFHL motif²¹. We suggest that association of Ifh1 and Fhl1 is mediated by interactions (direct or indirect) with Rap1 and/or the IFHL motif. Multiple Rap1 interactions with Fhl1 or Ifh1 minimize the need for the IFHL motif, whereas a protein–DNA interaction between Fhl1, Ifh1 and the IFHL motif minimizes the need for multiple interactions with Rap1. The precise location of Fhl1 and Ifh1 would depend on the quality of the IFHL motif and other sequences in the vicinity of the Rap1 site(s).

We investigated the role of Ifh1 in regulation of the RP genes using a strain in which the only copy of *IFH1* is under the control of the *GAL1* promoter. Galactose-induced expression of *IFH1* specifically increases transcription of several RP genes, but has little or no effect on two non-RP, Rap1-regulated genes (*ENO2* and *HYP2*), and *RPL18B*, whose promoter is not bound by Fhl1 or Ifh1 (Fig. 3a, b). Furthermore, upon galactose induction, Ifh1 associates with all transcriptionally activated promoters tested, whereas it does not associate with *ENO2*, *HYP2* and *RPL18B* (Fig. 3c). Rap1 association with RP and non-RP promoters was not altered significantly upon induction (Supplementary Fig. 2). Thus, Ifh1 is a transcriptional activator that is specific for RP genes that are targets of Ifh1 and Fhl1.

Fhl1 contains a forkhead-associated (FHA) domain that can be functionally replaced by an FHA domain from a transcriptional

activator from tobacco²². A hybrid protein containing the Gal4 DNA-binding domain and the FHA domain of Fhl1 activates transcription of *GAL7* (Fig. 3d). Furthermore, the Gal4–FHA domain fusion recruits Ifh1 to the *GAL7* promoter, whereas the Gal4 DNA binding domain alone or a Gal4 fusion to the Gcr2 activator does not (Fig. 3e). Thus, the FHA domain of Fhl1 is sufficient to recruit Ifh1 and activate transcription.

We examined the association of Rap1, Fhl1 and Ifh1 under three conditions of environmental stress: heat shock, osmotic shock and inhibition of the TOR pathway by rapamycin. As expected, these environmental stresses result in decreased RP gene expression (Fig. 4a). Interestingly, the three stress conditions cause a rapid dissociation of Ifh1, but not of Rap1 or Fhl1 association, at RP promoters (Fig. 4b–d) whereas the cellular level of Ifh1 does not decrease (Fig. 4e). The observation that Ifh1 is not required for association of Fhl1 strongly suggests that Rap1 interacts (directly or indirectly) with Fhl1. In addition, heat shock causes a large decrease in *GAL7* transcription (Fig. 3d) and Ifh1 recruitment (Fig. 3e; Supplementary Fig. 3) mediated by the hybrid protein containing the Gal4 DNA-binding domain and the FHA domain of Fhl1. Thus, in the presence of the Gal4–FHA fusion, the *GAL7* promoter responds to environmental stress indistinguishably from RP promoters.

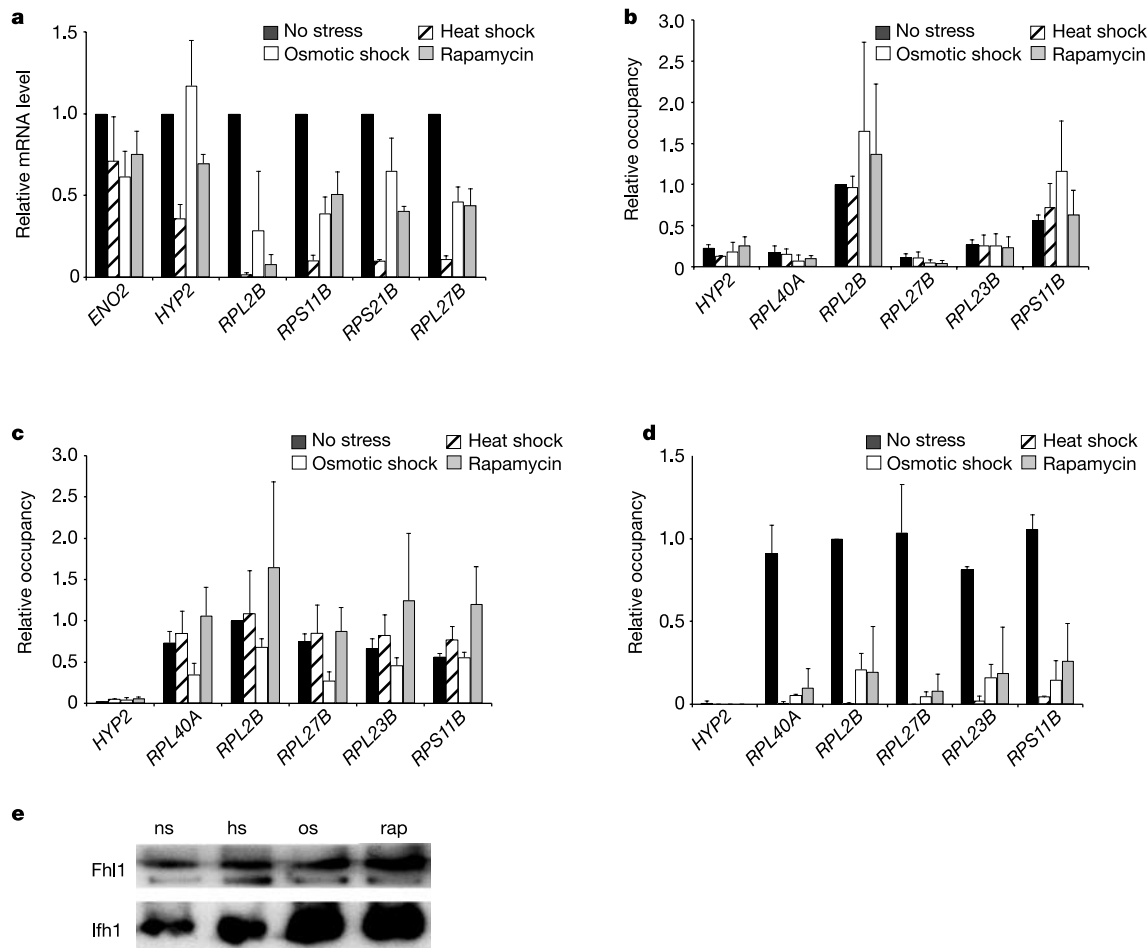


Figure 4 Effect of environmental stress on transcription and binding of Rap1, Fhl1 and Ifh1. **a**, RNA levels of indicated genes under various conditions relative to unstressed cells. **b**, Rap1 occupancy relative to the level at *RPL2B* in unstressed cells. **c**, Fhl1 occupancy relative to the level at *RPL2B* in unstressed cells. **d**, Ifh1 occupancy relative to the level at

RPL2B in unstressed cells. **e**, Western blot showing the cellular levels of Fhl1 and Ifh1 under conditions of no stress (ns), heat shock (hs), osmotic shock (os) and rapamycin treatment (rap). Equal amounts of cell extract were loaded in each lane. Error bars reflect the standard deviation of the mean.

Coordinate regulation of RP genes is of fundamental biological importance as it represents approximately 50% of all RNA polymerase II transcription, and it represents the primary mechanism by which protein synthesis and hence cell growth is regulated appropriately in response to optimal or stressful environmental conditions. Fhl1 and Ifh1 are the first proteins specifically linked to coordinate regulation of RP genes, and regulated association and dissociation of Ifh1 is a key step at which this control is exercised. Ifh1 dissociation, coupled with the short messenger RNA half-life of many RP mRNAs permits the rapid and efficient regulation of RP gene expression. However, Fhl1 and Ifh1 do not seem to associate with some RP promoters that are coordinately expressed, suggesting the existence of alternative regulatory mechanisms. Nevertheless, our results indicate that Fhl1 and particularly Ifh1 are key proteins that regulate most RP promoters in response to growth signals and environmental stress. □

Methods

Yeast strains and DNAs

Two strains were used for ChIP-chip analysis, both derived from BY4727 (ref. 23) with either Fhl1 (strain JTW1) or Ifh1 (strain JTW2) carboxy-terminally epitope-tagged with nine myc epitopes linked to a *TRP1* marker²⁴. JTW2 was also used for experiments involving Gal4 derivatives binding to the *GAL7* promoter. Gal4 derivatives were constructed in plasmid DH021, a derivative of YCplac111 (ref. 25) that contains the Gal4 DNA-binding domain (residues 1–147) amino-terminally tagged with three haemagglutinin (HA) epitopes, expressed from the *EFT2* promoter and upstream of the *CYC1* terminator. The Gal4 DNA-binding domain was fused either to residues 266–400 of Fhl1 (FHA

domain) or Gcr2. Derivates of *RPL20B* were constructed in plasmid HPIpV4 containing the *PGK1* core promoter (–246 to –7) upstream of the *HIS3* coding region¹⁰. The *RPL20B* upstream activating sequence (UAS) (–377 to –245) contains a Rap1 site and an IFHL motif (modified from CTAGGCCGCGG to CTAGGCCGAAG to facilitate cloning with a *SacII* restriction site). Wild-type Rap1 site: ACCCGTACA; mutated Rap1 site: AAAAGTACA; wild-type IFHL motif: CTAGGCCGAAG; and mutated IFHL motif: CTATTACTCGG. JTW3, which was used for galactose induction of *IFH1*, was created by sporulating a derivative of the diploid BY4743 (ref. 23) that contains one wild-type *IFH1* allele and one deleted *IFH1* allele, and also contains a derivative of plasmid YCplac33 (ref. 25) that has the *GAL1* promoter driving expression of *IFH1* N-terminally epitope-tagged with three HA epitopes (plasmid JZW1). JTW3 was selected as having *IFH1* expressed solely from plasmid JZW1 (note that this strain grows in medium containing glucose, presumably due to leaky expression of *IFH1*). JTW4, a derivative of JTW1 with Ifh1 C-terminally tagged with a tandem affinity purification (TAP) tag linked to a *URA3* marker, was used for all other experiments. For galactose induction, cells were grown at 30 °C in YP (yeast-peptone) media containing 2% raffinose to an absorbance of 0.7 at 600 nm ($A_{600}=0.7$), followed by addition of 2% galactose. For all other experiments cells were grown in YP media containing 2% glucose. For experiments involving environmental stress, cells were subjected to the following treatments: shift to 39 °C for 8 min; addition to 0.4 M NaCl for 12 min; addition to 100 nM rapamycin for 20 min.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was carried out with a modified version of a procedure described previously²⁶. Cells ($A_{600}=0.7$) were fixed in 1% formaldehyde for 20 min at room temperature, quenched for 5 min with glycine and lysed with zirconia-silica beads (BioSpec Products) in a mini-bead beater (BioSpec Products). Chromatin was first pelleted by centrifugation and then solubilized by sonication (Branson Sonifier 450, three times, 100% duty, power 5, 30 s for each cycle). Crosslinked chromatin was immunoprecipitated with protein A-sepharose beads (Amersham) and polyclonal antibody against Rap1, monoclonal antibodies against the HA (F7) and myc (9e10, all from Santa Cruz Biotechnology) epitopes, or with IgG-sepharose beads (Amersham). Quantitative polymerase chain reaction (PCR) analyses were performed in real time using

an Applied Biosystems 7700 sequence detector. Relative occupancy values were calculated by determining the apparent immunoprecipitation efficiency (amount of PCR product in the immunoprecipitated sample divided by the amount of PCR product in the input sample) and normalized to the level observed at the coding sequence of the *POL1* gene, which was defined as 1. This background binding was then subtracted to give a value in 'occupancy units'. Error bars shown reflect the standard deviation of the mean of independent experiments.

Western blotting

TAP (tandem affinity purification)-tagged Ifh1 was detected with peroxidase-anti-peroxidase antibody (Sigma) and Fhl1 was detected with anti-myc antibody (9e10, Santa Cruz Biotechnology).

Microarray analysis

Microarrays containing duplicate spots of 6,528 PCR products corresponding to nearly all yeast intergenic regions were hybridized with a mixture of amplified immunoprecipitate (labelled with Cy5 fluorescent dye) and input (labelled with Cy3 dye) samples, as described previously²⁷. Values shown are an average of two independent experiments. Conserved DNA motifs were identified using AlignACE²⁸. WebLogo was used to generate motif logos²⁹.

Transcriptional analysis

Total RNA was purified using Qiagen RNeasy columns with DNase I treatment. First-strand cDNA was synthesized using dT₁₆, and quantitative PCR in real time was performed on the resulting first-strand complementary DNA using primers specific to the gene of interest³⁰. RNA levels were determined relative to a control gene, *ACT1*.

Received 7 September; accepted 5 November 2004; doi:10.1038/nature03175.

- Mager, W. H. Control of ribosomal protein gene expression. *Biochim. Biophys. Acta* **949**, 1–15 (1988).
- Pogue-Geile, K. *et al.* Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA encoding the human S3 ribosomal protein. *Mol. Cell. Biol.* **11**, 3842–3849 (1991).
- Seshadri, T., Uzman, J. A., Oshima, J. & Campisi, J. Identification of a transcript that is down-regulated in senescent human fibroblasts. Cloning, sequence analysis, and regulation of the human L7 ribosomal protein gene. *J. Biol. Chem.* **268**, 18474–18480 (1993).
- Pierandrei-Amaldi, P. & Amaldi, F. Aspects of regulation of ribosomal protein synthesis in *Xenopus laevis*. *Genetica* **94**, 181–193 (1994).
- Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
- Lieb, J. D., Liu, X. L., Botstein, D. & Brown, P. O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genet.* **28**, 327–334 (2001).
- Moehle, C. M. & Hinnebusch, A. G. Association of RAP1 binding sites with stringent control of ribosomal protein gene transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 2723–2735 (1991).
- Klein, C. & Struhl, K. Protein kinase A mediates growth-regulated expression of yeast ribosomal protein genes by modulating RAP1 transcriptional activity. *Mol. Cell. Biol.* **14**, 1920–1928 (1994).
- Li, B., Nierras, C. R. & Warner, J. R. Transcriptional elements involved in the repression of ribosomal protein synthesis. *Mol. Cell. Biol.* **19**, 5393–5404 (1999).
- Mencia, M., Moqtaderi, Z., Geisberg, J. V., Kuras, L. & Struhl, K. Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol. Cell* **9**, 823–833 (2002).
- Shore, D. RAP1: a protean regulator in yeast. *Trends Genet.* **10**, 408–412 (1994).
- Morse, R. H. RAP, RAP, open up! New wrinkles for RAP1 in yeast. *Trends Genet.* **16**, 51–53 (2000).
- Neuman-Silberberg, F. S., Bhattacharya, S. & Broach, J. R. Nutrient availability and the *RAS*/cyclic AMP pathway both induce expression of the ribosomal protein genes in *Saccharomyces cerevisiae* but by different mechanisms. *Mol. Cell. Biol.* **15**, 3187–3196 (1995).
- Cardenas, M. E., Cutler, N. S., Lorenz, M. C., Di Como, C. J. & Heitman, J. The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev.* **13**, 3271–3279 (1999).
- Powers, T. & Walter, P. Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **10**, 987–1000 (1999).
- Lascaris, R. F., Mager, W. H. & Planta, R. J. DNA-binding requirements of the yeast protein Rap1p as selected *in silico* from ribosomal protein gene promoter sequences. *Bioinformatics* **15**, 267–277 (1999).
- Hermann-Le Denmat, S., Werner, M., Sentenac, A. & Thuriaux, P. Suppression of yeast RNA polymerase III mutations by FHL1, a gene coding for a fork head protein involved in rRNA processing. *Mol. Cell. Biol.* **14**, 2905–2913 (1994).
- Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- Cherel, I. & Thuriaux, P. The *IFH1* gene product interacts with a forkhead protein in *Saccharomyces cerevisiae*. *Yeast* **11**, 261–270 (1995).
- Goncalves, P. M. *et al.* Transcription activation of yeast ribosomal protein genes requires additional elements apart from binding sites for Abf1p or Rap1p. *Nucleic Acids Res.* **23**, 1475–1480 (1995).
- Beer, M. A. & Tavazoie, S. Predicting gene expression from sequence. *Cell* **117**, 185–198 (2004).
- Kim, M., Ahn, J. W., Song, K., Pack, K. H. & Pai, H. S. Forkhead-associated domains of the tobacco NtFHA1 transcription activator and the yeast Fhl1 forkhead transcription factor are functionally conserved. *J. Biol. Chem.* **277**, 38781–38790 (2002).
- Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1993).
- Cosma, M. P., Tanaka, T. & Nasmyth, K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299–311 (1999).
- Gietz, R. D. & Sugino, A. New yeast-*Escherichia coli* shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. *Gene* **74**, 527–534 (1988).
- Aparicio, O. M., Geisberg, J. V. & Struhl, K. In *Current Protocols in Molecular Biology* (eds Ausubel, F. A. *et al.*) 21.3.1–21.3.17 (John Wiley & Sons, New York, 2004).
- Moqtaderi, Z. & Struhl, K. Genome-wide occupancy of the RNA polymerase III machinery in

- Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol. Cell. Biol.* **24**, 4118–4127 (2004).
- Roth, F. P., Hughes, J. D., Estep, P. W. & Church, G. M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnol.* **16**, 939–945 (1998).
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
- Reid, J. L., Moqtaderi, Z. & Struhl, K. Eaf3 regulates the global pattern of histone acetylation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **24**, 757–764 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Shore for communicating similar results before publication, N. Reppas for help with computational analysis, J. Weiner for assistance with microarrays, and F. Bachand, C. Baisden, and Z. Moqtaderi for technical assistance. We thank J. Geisberg, P. Mason and Z. Moqtaderi for discussions. J.T.W. was supported by a long-term EMBO fellowship, and D.B.H. was supported by a Helen Hay Whitney postdoctoral fellowship. This work was supported by grants to K.S. from the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S. (kevin@hms.harvard.edu).

Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1

Stephan B. Schawalter¹, Mehdi Kabani², Isabelle Howald¹, Urmila Choudhury¹, Michel Werner² & David Shore¹

¹Department of Molecular Biology and NCCR Program 'Frontiers in Genetics' University of Geneva, Sciences III, 30, quai Ernest-Ansermet, Geneva 4, CH-1211, Switzerland

²Service de Biochimie et Génétique Moléculaire, CEA Saclay, 91191 Gif-Sur-Yvette, Cedex, France

Regulation of ribosome biogenesis is central to the control of cell growth¹. In rapidly growing yeast cells, ribosomal protein (RP) genes account for approximately one-half of all polymerase II transcription-initiation events¹, yet these genes are markedly and coordinately downregulated in response to a number of environmental stress conditions^{2–4}, or during the transition from fermentation to respiration⁵. Although several conserved signalling pathways (TOR, RAS/protein kinase A and protein kinase C) impinge upon RP gene transcription¹, little is known about how initiation at these genes is controlled. Rap1 (refs 6, 7) and more recently Fhl1 (ref. 8) were shown to bind upstream of many RP genes. Here we show that the essential protein Ifh1 binds to and activates many RP gene promoters under optimal growth conditions in *Saccharomyces cerevisiae*. Ifh1 is recruited to RP gene promoters through the forkhead-associated domain of Fhl1. Ifh1 binding decreases when RP genes are downregulated either by TOR inhibition or nutrient depletion, and is restored after release from starvation or upon regulated induction of *IFH1* expression. These findings indicate a central role for Ifh1 and Fhl1 in RP gene regulation.

A recent global chromatin location analysis in yeast uncovered the forkhead-like protein Fhl1 as a factor highly specific for RP gene promoters⁸. Fhl1 was originally identified as a gene dosage suppressor of RNA polymerase (Pol) III mutations required for normal growth⁹. The slow growth of *fhl1Δ* cells is suppressed by elevated gene dosages of *IFH1*, suggesting that both Fhl1 and Ifh1 are involved in ribosomal RNA transcription or maturation¹⁰. In an attempt to resolve these apparently conflicting findings, we first

tested the prediction⁸ that Fhl1 is required for normal expression of the Pol-II-driven RP genes. We examined two representative genes (*RPL9A* and *RPL30*) in *fhl1Δ* cells and found reduced levels of both transcripts relative to actin (*ACT1*), which does not bind Fhl1 (ref. 8; see also Supplementary Table 1), suggesting that Fhl1 binding to RP gene upstream regulatory sequences (URS) is required for full expression (Fig. 1a). Similar results were seen in *fhl1Δ ifh1Δ* cells (Fig. 1a), and may explain why both *fhl1Δ* and *fhl1Δ ifh1Δ* strains grow extremely slowly compared with wild type (ref. 10 and data not shown).

We next used chromatin immunoprecipitation (ChIP) to test the idea that the essential Ifh1 protein is also involved in RP gene transcription, and observed specific binding to three different RP gene URS (*RPL30*, *RPL37A* and *RPL9*) under optimal growth conditions (Fig. 1b). Similarly, we confirmed that both Rap1 (ref. 6) and Fhl1 (ref. 8) also bind to these URS (Fig. 1b, c). Notably, we found that Ifh1 binding to these genes is undetectable in *fhl1Δ* cells (Fig. 1b), suggesting that Fhl1 has an essential role in the recruitment of Ifh1 to RP genes. This effect is specific to Ifh1, because neither *fhl1Δ* nor the *fhl1Δ ifh1Δ* double mutant shows any loss of Rap1 binding (Fig. 1c). Conversely, a non-RP gene bound by Rap1 (*PGK1*) shows no binding of either Fhl1 or Ifh1 (Fig. 1c), indicating that Rap1 binding itself does not lead to either Fhl1 or Ifh1 promoter association.

As an exponentially growing yeast culture begins to deplete glucose from the medium, cells initiate a global re-programming of transcription in preparation for the metabolic transition from fermentation to respiration¹¹. Well before this 'diauxic shift', a marked drop in RP gene transcript levels is observed⁵, the mechanism of which is unknown. To address this issue we used ChIP to measure association of Rap1, Fhl1 and Ifh1 at three different RP gene promoters as cells prepared for the diauxic shift. We observed a

marked decline in Ifh1 promoter binding at *RPL30* (and the two other RP genes tested; data not shown) that coincided with, or immediately preceded, the observed drop in messenger RNA level (Fig. 1d). In contrast, promoter association of Rap1 and Fhl1 remains constant during this period. The protein levels of all three factors are largely unchanged during the period of RP gene repression (Supplementary Fig. S1a). These results establish a tight temporal correlation between loss of Ifh1 promoter association and downregulation of RP gene transcription in a wild-type setting, and suggest that controlled promoter association of Ifh1 may be a key regulatory mechanism for RP genes. We also examined the RP gene association of the three proteins as stationary phase cells resume exponential growth after inoculation into fresh medium. As expected, *RPL30* transcripts accumulate rapidly (approximately sevenfold increase by 15 min; Supplementary Fig. S1b). Notably, at this early time point we observed a roughly threefold increase in Ifh1 URS association, whereas both Fhl1 and Rap1 association remained constant. These data reinforce the notion that Ifh1 binding at RP genes is associated with (and possibly required for) full activation of these genes (see below).

To investigate whether reversible Ifh1 binding is more generally associated with RP gene regulation, we examined the response to rapamycin treatment, which inhibits TOR, a highly conserved phosphatidylinositol-kinase-like growth regulator in yeast¹². As previously reported^{13–15}, we found that rapamycin causes a rapid drop in the mRNA levels of RP genes (Fig. 1e). Notably, Ifh1 association with three different RP gene promoters begins to drop by the earliest time point tested (5 min after rapamycin addition), and is tenfold lower compared with untreated cells by 30 min. This drop in Ifh1 promoter association precedes any measurable change in RP gene mRNA levels (perhaps by as much as 10 min), suggesting that it may be a cause of transcriptional downregulation. Notably,

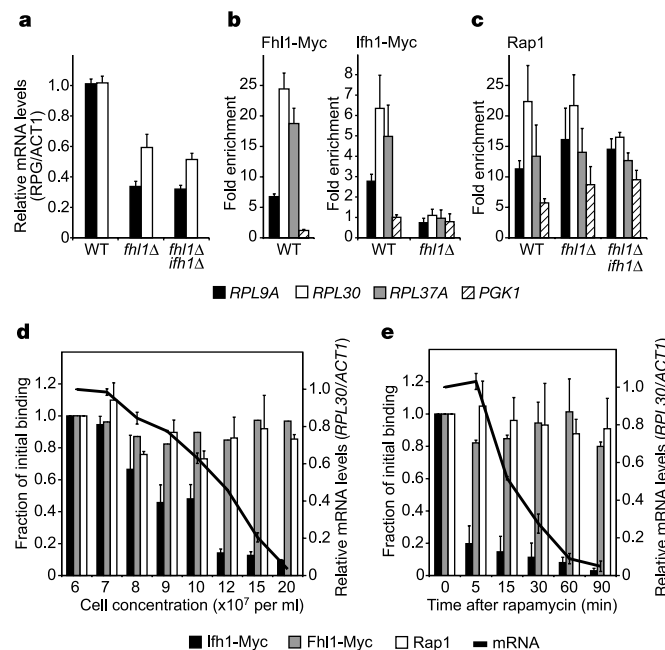


Figure 1 Ifh1, Fhl1 and Rap1 binding at RP gene (RPG) promoters, and mRNA levels during logarithmic growth, before the diauxic shift or after TOR inhibition. **a**, *RPL9A* and *RPL30* transcript levels relative to actin in wild-type (WT), *fhl1Δ* and *fhl1Δ ifh1Δ* strains (logarithmic growth: 2×10^7 cells ml⁻¹). **b**, Promoter association of Fhl1- and Ifh1-Myc at three RP gene promoters and the glycolytic gene *PGK1* in logarithmic wild-type and *fhl1Δ* strains. Mean values and standard deviations (s.d.) of three independent experiments are shown. **c**, Binding of Rap1 in wild-type, *fhl1Δ* and *fhl1Δ ifh1Δ* strains, as in **b**. **d**, Diauxic shift. Fhl1- or Ifh1-Myc strains were processed for ChIP and mRNA

analysis at the indicated cell concentrations (bottom). Binding for each factor at the initial time point (fold enrichment; see **b**, **c**) was set to 1. Three independent ChIP reactions and mRNA measurements for each point were performed (mean values and s.d. for one such experiment are shown). **e**, TOR inhibition. Fhl1- or Ifh1-Myc strains were grown to 2×10^7 per ml, when rapamycin was added to 200 ng ml⁻¹ ($t = 0$). Preparation and analysis of samples was as in **d**, and data shown are the average and s.d. of three independent experiments.

and like the diauxic shift, neither Rap1 nor Fhl1 promoter binding appear to change after TOR inhibition, and total Ifh1 protein levels are unchanged at the time that the protein disappears from the RP gene promoters (Supplementary Fig. S1a).

To determine whether Fhl1 is directly involved in recruiting Ifh1 to RP genes, we fused parts of Fhl1 to the Gal4 DNA-binding domain (GBD). We then asked whether tethering of these GBD–Fhl1 hybrids (Fig. 2a) to a promoter containing Gal4-binding sites, where Fhl1 is normally not bound, would be sufficient to recruit Ifh1 there. ChIP analysis showed that this is indeed the case (Fig. 2b). Ifh1 binding at the *GAL7* promoter in cells expressing GBD–Fhl1 hybrids is quantitatively similar to that observed at native RP gene promoters (data not shown), suggesting that Fhl1 is sufficient for robust Ifh1 recruitment at an ectopic site. Significantly, we found that the forkhead-associated (FHA) domain of Fhl1, and perhaps a small amount of flanking sequence, is sufficient for Ifh1 recruitment, whereas all hybrids tested that lack this domain fail to recruit Ifh1. Furthermore, GBD–Fhl1-mediated Ifh1 recruitment is sensitive to rapamycin, suggesting that a primary downstream effect of TOR inhibition may be to interfere with the ability of Ifh1 to associate with Fhl1 at RP gene promoters. Finally, although recruitment of Ifh1 is consistently associated with activation of the *GAL7-lacZ* reporter (and lack of recruitment is associated with no activation) there is no strict quantitative relationship between the amount of Ifh1 binding and the extent of activation (Fig. 2c). The carboxy terminus of Fhl1, and to a lesser extent the FH domain, appear to have a repressive function in this context.

To ask whether the Fhl1–Ifh1 interactions described above at three RP gene promoters are common to most, or all, RP genes, we conducted a chromosomal binding analysis of both Ifh1 and Fhl1 through hybridization of immunoprecipitated DNA to yeast intergenic microarrays^{6,16,17}. As reported previously⁸, we found that Fhl1 chromatin binding *in vivo* is highly specific for RP genes. We found the same to be true for Ifh1, the chromosomal binding sites of which remarkably overlap those found for Fhl1 (Supplementary Fig. S2a, b). Fhl1 and Ifh1 bound to 53% (72) and to 45% (61) ($P < 0.001$) of the 137 RP gene promoters, respectively. Moreover, 34% (46) of RP gene promoters were bound by both factors. Owing to the stringency of our analysis and technical limitations, we believe that these data actually underestimate the co-localization of these factors. We repeated our analysis with cells that had been

treated with rapamycin before crosslinking, and found a markedly different result for Ifh1 binding, as predicted by our study of three specific RP genes. Specifically, rapamycin treatment reduced binding to background levels at all but one RP gene, and no new sites of Ifh1 binding were identified (Supplementary Fig. S2 and Supplementary Table 1). As expected from our initial studies, we observed very little effect of rapamycin on global Fhl1 binding.

We next asked whether Ifh1 functions as a direct transcriptional activator of native RP genes, as the above results strongly imply. To address this question we constructed a strain in which the endogenous *IFH1* gene is controlled by the galactose-inducible *GAL1* promoter. This strain grows slowly and contains very low levels of Ifh1 protein on medium containing glucose (which represses

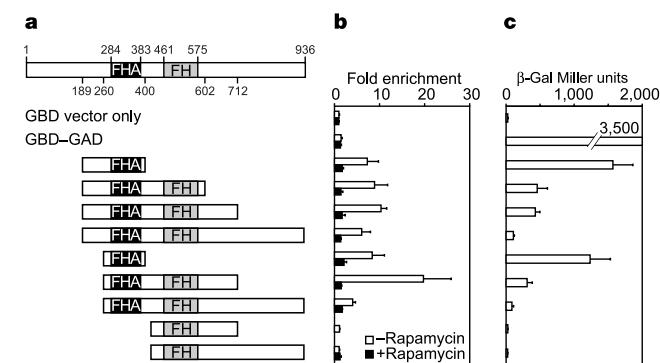


Figure 2 Rapamycin-sensitive recruitment of Ifh1 to an ectopic site (*GAL7*) via the FHA domain of GBD–Fhl1 hybrids. **a**, Schematic representation of native Fhl1 (ref. 28) and GBD–Fhl1 hybrid proteins. Numbers above the bar indicate amino acid positions of domain boundaries; those below indicate the fusion endpoints of GBD–Fhl1 hybrids. **b**, Ifh1–Myc binding at *GAL7* as measured by ChIP in cells growing exponentially or 90 min after rapamycin treatment. Cells express control (GDB alone or GBD–GAD) or GBD–Fhl1 hybrid proteins, as indicated in **a**. The mean values and s.d. of three independent experiments are shown. **c**, β -Galactosidase activity in cells expressing the proteins indicated in **a**.

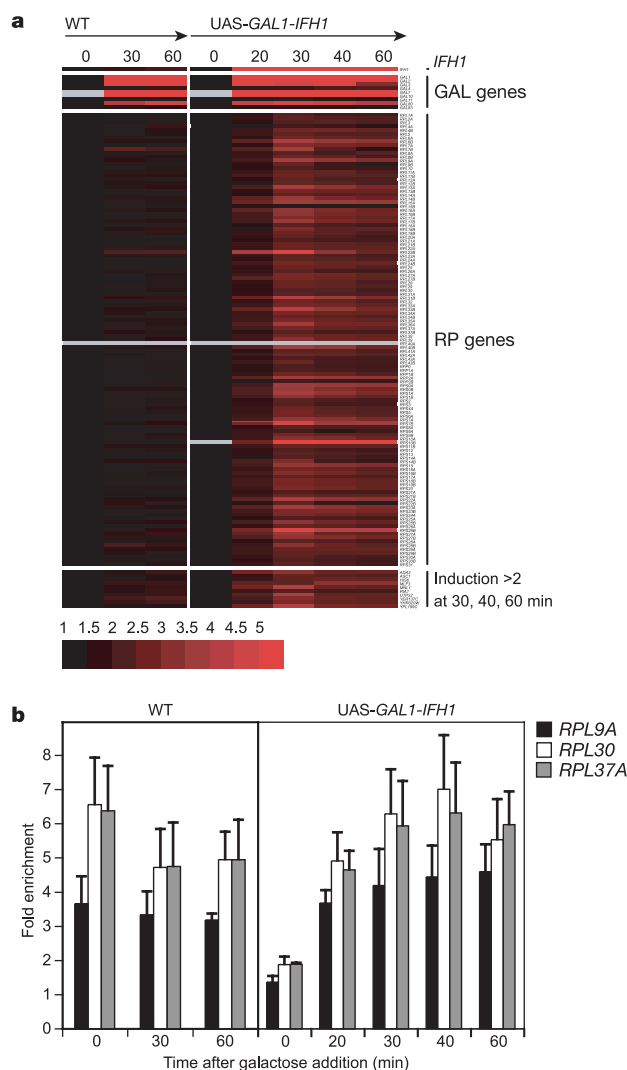


Figure 3 Ifh1 directly activates RP gene transcription. Isogenic wild-type and UAS-*GAL1-IFH1* cells were grown to 2×10^7 cells ml^{-1} in YPLG medium (containing lactic acid and glycerol) and $t = 0$ samples collected. Galactose was then added to 2% and samples collected at the indicated time points (in minutes). **a**, mRNA expression ratios of induced genes relative to the initial YPLG culture ($t = 0$), as measured by microarray analysis, are shown for each time point. Genes specifically induced in UAS-*GAL1-IFH1* cells (>2 -fold) at $t = 30$, 40 and 60 were grouped. Red indicates >1 -fold increase (induction) and grey signifies that no reliable measurement was obtained. Functional classes of genes are indicated, except from the ten genes grouped at the bottom (see text and Supplementary Fig. S3 for details). **b**, Ifh1–Myc binding to three RP gene promoters (*RPL9A*, *RPL30* and *RPL37A*) for each time point as measured by ChIP. The mean values and s.d. of three independent experiments are shown. See Supplementary Information for further details.

UAS-*GAL1-IFH1*), but grows normally on galactose (inducing) medium (data not shown). We examined, by genome-wide microarray analysis, the effect of inducing *IFH1* expression in this strain by shifting cells from non-inducing (lactic acid and glycerol) to inducing medium. Nearly all mRNAs of RP genes were detectably upregulated only 20 min after galactose addition, and by 30 min were typically two- to fivefold induced, in all cases exceeding the modest induction observed for some RP genes in the control (*IFH1* wild type) strain (Fig. 3a). In the three cases examined directly by ChIP, this increase in RP gene transcript was associated with increased *Ihf1* promoter binding, again unlike the control strain where *Ihf1* binding remained constant (Fig. 3b). Because the RP genes were among the primary targets after induction of *Ihf1*, we conclude that *Ihf1* is a direct transcriptional activator of these genes. Only ten other Pol II protein-encoding genes were induced as rapidly and as strongly as the RP genes (Fig. 3a). Among these genes *UTP22*, *ASC1* and *RIA1* are bound by both *Rap1* and *Fhl1* (ref. 8; see also Supplementary Fig. S2) and are involved in ribosome biogenesis. Our data thus point to a possible role for *Ihf1* in coordinating RP gene activation with other key steps in ribosome biogenesis.

The results reported here support a simple model in which normal RP gene expression under favourable growth conditions requires recruitment of the essential (co)activator *Ihf1* to RP gene promoters through an interaction with the FHA domain of promoter-bound *Fhl1*. This model shares several features with that proposed to explain the action of two other FH proteins, *Fkh1* and *Fkh2*, together with the essential protein *Ndd1* in activation of the 'CLB2 cluster' of G2/M-specific genes^{18,19}. One curious property of both systems is that the requirement of the co-activator for cell growth (*Ihf1* here; *Ndd1* for the *Fkh1/2* system) can be bypassed by deletion of the DNA-binding recruitment factor¹⁰. In cells lacking *Ihf1*, *Fhl1* may repress RP gene transcription, either alone or with a co-repressor, and thus block cell growth. In the absence of both proteins, activation of RP genes is low but sufficient to permit cell growth.

Our data also imply that the *Fhl1*–*Ihf1* interaction constitutes a major regulatory node in the system: both glucose (possibly through the RAS/protein kinase A pathway) and TOR activity are required for full recruitment of *Ihf1* to RP gene promoters. Significantly, inactivation of either one of these two major growth stimulatory pathways results in a rapid loss of *Ihf1* promoter binding, which is followed by a large drop in RP gene mRNA levels. The mechanism(s) by which this occurs are unknown. The fact that *Ihf1* is recruited by the FHA domain of *Fhl1* (Fig. 2b), and that FHA domains in various other proteins are phospho-peptide-binding motifs²⁰, suggests that *Ihf1* recruitment at RP gene promoters may be sustained by active, reversible phosphorylation of a putative *Fhl1*-binding domain in *Ihf1*. Whether this occurs through the direct action of protein kinase A, TOR, or both kinases, is at present unknown.

The findings reported here do not exclude the possibility that other important control mechanisms act directly at RP gene promoters. Indeed, the coordinated downregulation of RP genes under a wide range of stress conditions, and the importance to the cell of this regulation from an energetic standpoint¹, suggests that multiple mechanisms are probably involved. An obvious additional candidate for regulatory input at RP gene promoters is *Rap1* (ref. 7, 21), which could function in promoting proper *Fhl1* binding, or in the recruitment of the *Esa1* histone acetyltransferase and/or the TAF-containing form of TFIID^{22–24}. Finally, the recently identified growth regulator *Sfp1* is an additional candidate regulator of RP genes^{25–27}. An understanding of RP gene regulation will provide a detailed picture of how the yeast cell integrates various growth signals at the level of the promoter. This knowledge should help to

dissect further the signal pathways that regulate ribosome biogenesis and growth in yeast, and is likely to have important implications in more complex multicellular organisms. □

Received 30 September; accepted 18 November 2004; doi:10.1038/nature03200.

- Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
- Gasch, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
- Causton, H. C. et al. Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**, 323–337 (2001).
- Rep, M., Krantz, M., Thevelein, J. M. & Hohmann, S. The transcriptional response of *Saccharomyces cerevisiae* to osmotic shock. *Hot1p* and *Msn2p/Msn4p* are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *J. Biol. Chem.* **275**, 8290–8300 (2000).
- Ju, Q. & Warner, J. R. Ribosome synthesis during the growth cycle of *Saccharomyces cerevisiae*. *Yeast* **10**, 151–157 (1994).
- Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. Promoter-specific binding of *Rap1* revealed by genome-wide maps of protein–DNA association. *Nature Genet.* **28**, 327–334 (2001).
- Shore, D. *RAP1*: a protean regulator in yeast. *Trends Genet.* **10**, 408–412 (1994).
- Lee, T. I. et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- Hermann-Le Denmat, S., Werner, M., Sentenac, A. & Thuriaux, P. Suppression of yeast RNA polymerase III mutations by *FHL1*, a gene coding for a fork head protein involved in rRNA processing. *Mol. Cell. Biol.* **14**, 2905–2913 (1994).
- Cherel, I. & Thuriaux, P. The *IFH1* gene product interacts with a fork head protein in *Saccharomyces cerevisiae*. *Yeast* **11**, 261–270 (1995).
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
- Jacinto, E. & Hall, M. N. Tor signalling in bugs, brain and brawn. *Nature Rev. Mol. Cell Biol.* **4**, 117–126 (2003).
- Cardenas, M. E., Cutler, N. S., Lorenz, M. C., Di Como, C. J. & Heitman, J. The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev.* **13**, 3271–3279 (1999).
- Hardwick, J. S., Kuruvilla, F. G., Tong, J. K., Shamji, A. F. & Schreiber, S. L. Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc. Natl Acad. Sci. USA* **96**, 14866–14870 (1999).
- Powers, T. & Walter, P. Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **10**, 987–1000 (1999).
- Ren, B. et al. Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- Harismendy, O. et al. Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J.* **22**, 4738–4747 (2003).
- Koranda, M., Schleiffer, A., Endler, L. & Ammerer, G. Forkhead-like transcription factors recruit *Ndd1* to the chromatin of G2/M-specific promoters. *Nature* **406**, 94–98 (2000).
- Jorgensen, P. & Tyers, M. The fork'ed path to mitosis. *Genome Biol.* **1**, R1022.1–1022.4 (2000).
- Durocher, D. & Jackson, S. P. The FHA domain. *FEBS Lett.* **513**, 58–66 (2002).
- Klein, C. & Struhl, K. Protein kinase A mediates growth-regulated expression of yeast ribosomal protein genes by modulating *RAP1* transcriptional activity. *Mol. Cell. Biol.* **14**, 1920–1928 (1994).
- Mencia, M., Moqtaderi, Z., Geisberg, J. V., Kuras, L. & Struhl, K. Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol. Cell* **9**, 823–833 (2002).
- Reid, J. L., Iyer, V. R., Brown, P. O. & Struhl, K. Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of *Esa1* histone acetylase. *Mol. Cell* **6**, 1297–1307 (2000).
- Rohde, J. R. & Cardenas, M. E. The tor pathway regulates gene expression by linking nutrient sensing to histone acetylation. *Mol. Cell. Biol.* **23**, 629–635 (2003).
- Jorgensen, P., Nishikawa, J. L., Breitkreutz, B. J. & Tyers, M. Systematic identification of pathways that couple cell growth and division in yeast. *Science* **297**, 395–400 (2002).
- Marion, R. M. et al. *Sfp1* is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression. *Proc. Natl Acad. Sci. USA* **101**, 14315–14322 (2004).
- Jorgensen, P. et al. A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes Dev.* **18**, 2491–2505 (2004).
- Kim, M., Ahn, J. W., Song, K., Paek, K. H. & Pai, H. S. Forkhead-associated domains of the tobacco NtFHA1 transcription activator and the yeast *Fhl1* forkhead transcription factor are functionally conserved. *J. Biol. Chem.* **277**, 38781–38790 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Warner, E. Di Mauro, K. Mizuta, G. Ammerer, P. Thuriaux and M. Tyers for discussions; J. Warner, K. Struhl and M. Tyers for communicating unpublished results; A. Bianchi for comments on the manuscript; members of the Shore laboratory for discussions; P. Descombes, C. Barraclough, C. Delucinge, O. Schaad, P. Soularue and X. Gidrol for help with microarray experiments and analysis; S. Gasser for providing antibodies; and N. Roggli for artwork. This work was supported by the Human Frontier Science Program, the Swiss National Science Foundation, the NCCR program 'Frontiers in Genetics' (Swiss National Science Foundation), and the Canton of Geneva.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.S. (David.Shore@molbio.unige.ch).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Andy Douglas (4975)
Nils Moeller (4953)
Irene Vigila-Alton (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Advertising Production

Manager: Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Niamh Shields

European Satellite Office

**Germany/ Austria/ Italy/
The Netherlands/ Belgium:**

Patrick Phelan
Tel + 49 89 54 90 57 11
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
Sharon de Weert
Tel + 44 (0) 20 7843 4970
e-mail: s.deveert@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Planning ahead

It may be possible to leave buying holiday gifts until the last minute, but it might not be prudent to procrastinate when shopping for conferences and workshops. This is especially true for postdocs and graduate students, and doubly so for those choosing their first meeting. Two factors make it important to plan ahead: availability and funding.

Availability, or getting in, may be relatively easy for larger society-based conferences — but such meetings can be the least beneficial to younger scientists. Although many will feature career workshops or talks, access to plenary speakers is often difficult as they tend to get mobbed after their presentations.

At smaller conferences, it is easier to network with a senior scientist whose work meshes with your own career — although these meetings usually have a restricted number of places. And the chance to learn key skills at workshops can be pivotal early in your career — but many of the most popular workshops are inevitably oversubscribed. The best way to guarantee a slot is to research the options in our Events Directory published in this issue, making your choice and registering straight away.

Making sure someone pays or reimburses you before or after you register is another matter entirely. If you are presenting a poster or a talk but aren't funded, talk to your principal investigator. If they can't pay, your department may have money. Some professional societies also offer travel grants, as do some funding bodies. Sometimes you have to be a confirmed participant in a workshop before you can apply for the funds; other schemes award funds on a first-come, first-served basis.

Whatever the case, as with Christmas shopping, it pays to plan ahead — and to be armed with a credit card to pay the bills.

Paul Smaglik
Naturejobs editor



WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS